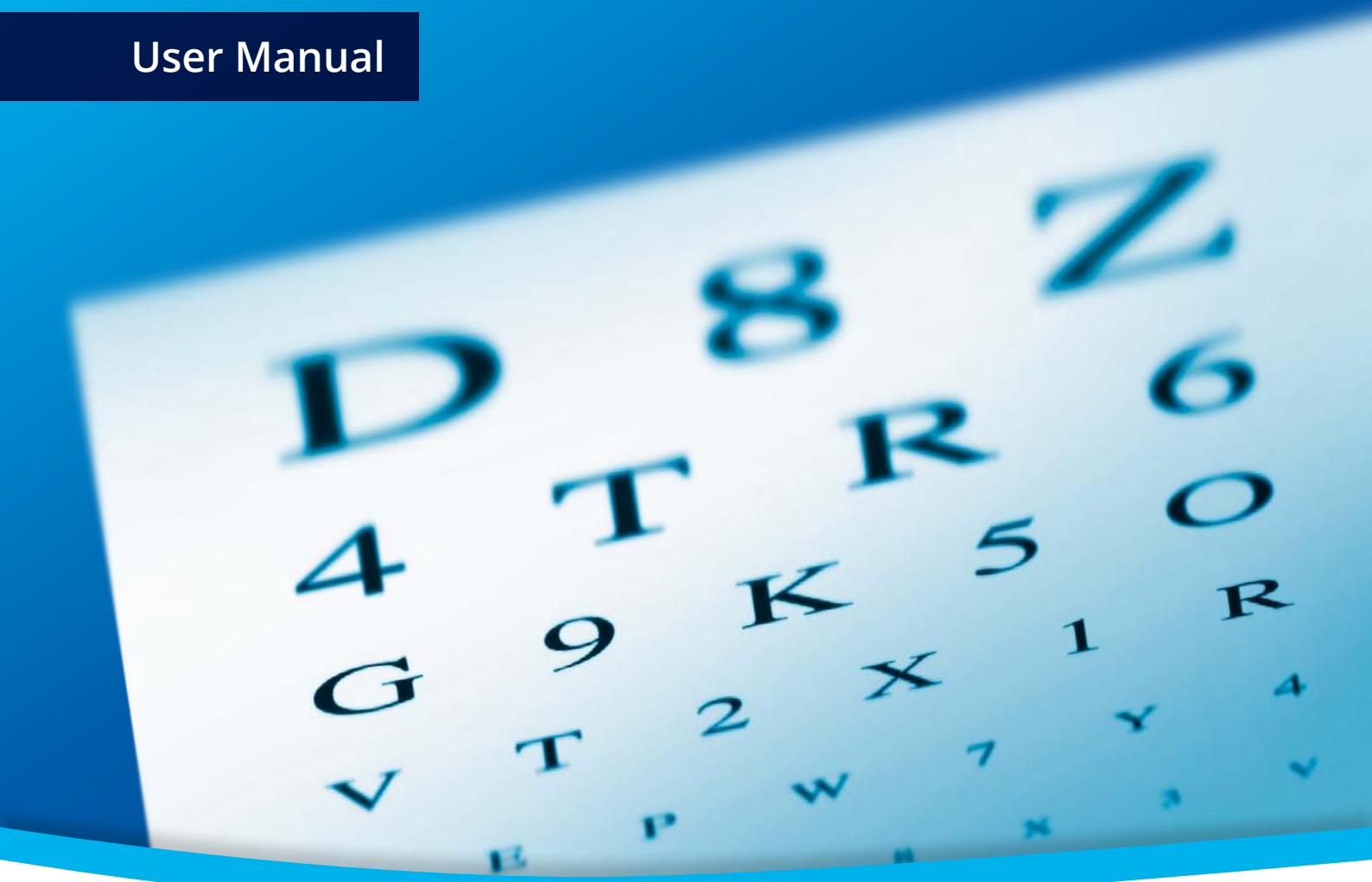
User Manual



3-Heights™ OCR Enterprise Add-On

Version 4.5

PDF-TOOLS.COM
Premium PDF Technology

Contents

1 Introduction . . . 1
1.1 Overview . . . 1
1.2 Supported OCR Engines . . . 1
1.3 Operating Systems . . . 2
2 Installation and Configuration . . . 2
2.1 ABBYY . . . 3
2.2 OCR Service . . . 6
2.3 Tesseract for Windows . . . 7
2.4 Tesseract for Unix . . . 7
2.5 Installing Products Using the Add-On . . . 8
3 Licensing . . . 9
3.1 ABBYY FineReader Engine . . . 9
4 Reference Manual . . . 10
4.1 Architecture . . . 10
4.2 OCR Related Switches and Functions . . . 12
4.3 Examples . . . 12
4.4 Impacts of Using OCR . . . 14
5 Configuration and Performance . . . 15
5.1 Configuration for ABBYY . . . 15
5.2 Configuration for Tesseract . . . 15
5.3 Performance Overview . . . 15
5.4 Performance Figures . . . 16
6 Licensing and Copyright . . . 18
7 Contact . . . 18
A Appendix ABBYY FineReader 8 . . . 18
A.1 Languages . . . 19
A.2 OCR Engine Parameters . . . 21
B Appendix ABBYY FineReader 10 . . . 29
B.1 Profiles for ABBYY 10 . . . 29
B.2 ABBYY 10 Parameters . . . 30
C Appendix ABBYY FineReader 11 . . . 41
C.1 Languages for ABBYY 11 . . . 41
Natural languages . . . 41
Technical . . . 43
C.2 Profiles for ABBYY 11 . . . 44
Predefined Profiles . . . 44
Custom Profiles . . . 45
Profile Parameters . . . 45
D Appendix Tesseract . . . 68
D.1 Languages . . . 68
D.2 OCR Engine Parameters . . . 68

1 Introduction

1.1 Overview

The 3-Heights™ OCR Enterprise Add-On adds optical character recognition (OCR) to several 3-Heights™ products. Documents generated with a product that makes use of the 3-Heights™ OCR Enterprise Add-On contain searchable texts.

The optional 3-Heights™ OCR Enterprise Add-On recognizes text content and embeds this text as Unicode text in a PDF or PDF/A file. Thanks to this process, the PDF files are searchable and the texts can be used in other applications, databases etc.

The add-on can be used in combination with one of the following products:

- 3-Heights™ Image to PDF Converter
- 3-Heights™ PDF to PDF/A Converter
- 3-Heights™ Document Converter

1.2 Supported OCR Engines

Each OCR engine is interfaced via a corresponding plug-in. Plug-ins are provided by PDF Tools. Currently the following plug-ins are available:

ABBYY FineReader OCR Engine 8

This plug-in is based on the ABBYY FineReader OCR Engine 8 and has – in its enterprise version – no page limitations. It can be used on one client or server computer with 1 or 2 CPUs. ABBYY 8 licenses are no longer sold.

Supported OS: Windows 2000 and later

ABBYY FineReader OCR Engine 10

This plug-in is based on the ABBYY FineReader OCR Engine 10 and comes in different throughput models.

ABBYY FineReader OCR Engine 11

This plug-in is based on the ABBYY FineReader OCR Engine 11 and comes in different throughput models.

OCR Service (using ABBYY)

This plug-in implements an OCR service that uses the Abbyy FineReader OCR Engine. The OCR service should be used in any situation where files are processed in parallel (for example with the Document Converter).

It is available in version 1.9.6.1 or later.

Supported OS: Windows XP/2003 or later

Tesseract

This plug-in interfaces the Tesseract OCR engine.

It is available in version 1.91.18.0 or later.

Supported OS: Windows 2000 or later, Linux

1.3 Operating Systems

The 3-Heights™ OCR Enterprise Add-On is available for the following operating systems:

- Windows XP, Vista, 7, 8, 8.1 - 32 and 64 bit
- Windows Server 2003, 2008, 2008-R2, 2012, 2012-R2 - 32 and 64 bit
- Linux (32 and 64 bit)

2 Installation and Configuration

This chapter describes:

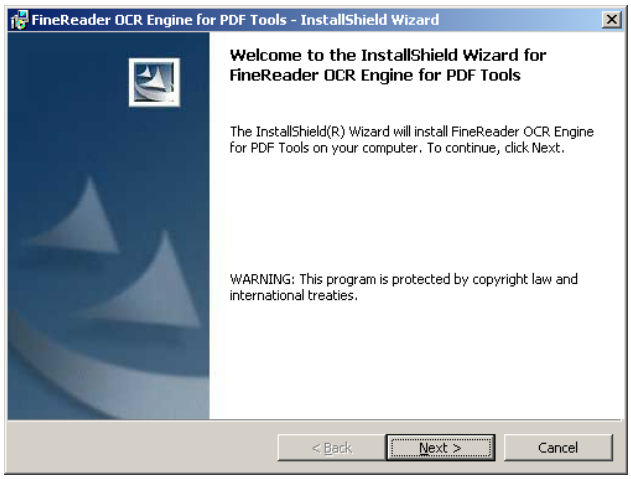
- How to install the ABBYY FineReader OCR Engine and how to activate its license.
- How to install the 3-Heights™ PDF OCR Service.
- How to install the Tesseract OCR Engine.
- How to setup a tool that uses the 3-Heights™ OCR Enterprise Add-On so that is uses the OCR engine directly or via OCR service.

2.1 ABBYY

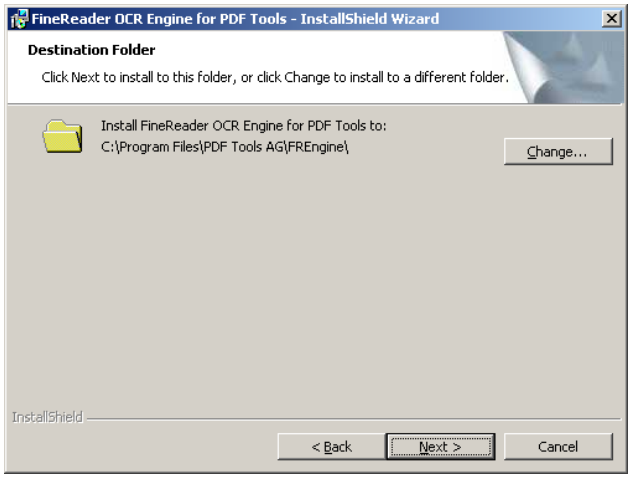
Installation of ABBYY OCR FineReader Engine

The Windows version of the 3-Heights™ OCR Enterprise Add-On is distributed in MSI-format (Microsoft Software Installation). The MSI file is named FineReader OCR Engine for PDF Tools.msi.
The screenshots are from ABBYY 8, the installation for ABBYY 10 and ABBYY 11 is equivalent.

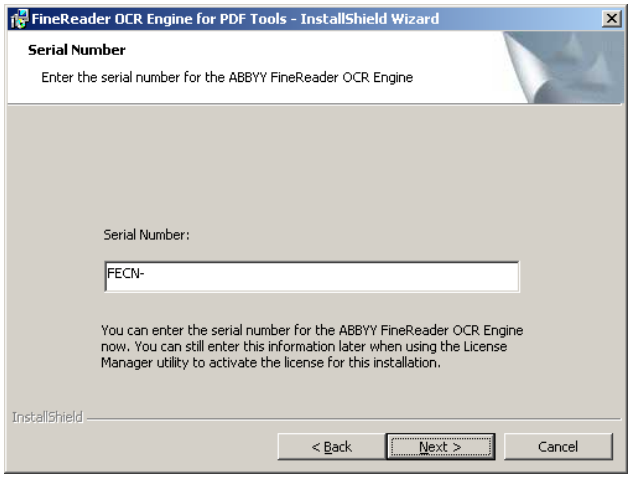
1. Download the ZIP archive "OCRE<version><platform>.zip" (e.g. OCRE190Win32.zip) from your download account at www.pdf-tools.com.
2. Extract the file *FineReader OCR Engine for PDF Tools.msi* from the ZIP archive.
3. Double-click the MSI file to start the installation wizard.
4. In the first dialog box of the installation wizard press the button "Next".


FineReader OCR Engine for PDF Tools - InstallShield Wizard
Welcome to the InstallShield Wizard for FineReader OCR Engine for PDF Tools
The InstallShield(R) Wizard will install FineReader OCR Engine for PDF Tools on your computer. To continue, click Next.
WARNING: This program is protected by copyright law and international treaties.
< Back
Next >
Cancel

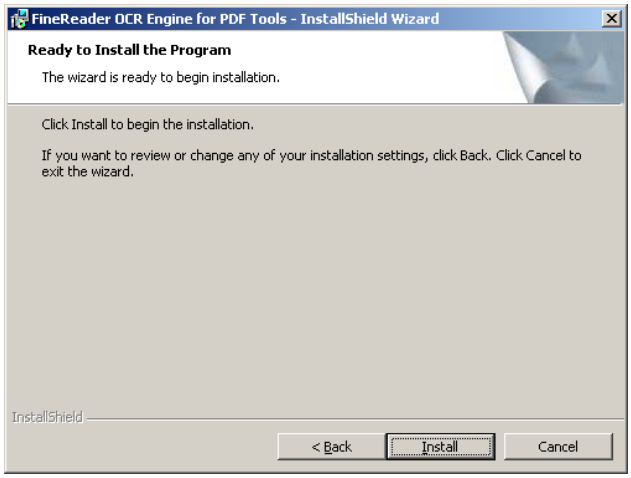
5. Adjust the destination folder if required using the button "Change..." Press "Next" when done.


FineReader OCR Engine for PDF Tools - InstallShield Wizard
Destination Folder
Click Next to install to this folder, or click Change to install to a different folder.
Install FineReader OCR Engine for PDF Tools to:
C:\Program Files\PDF Tools AG\FREngine\
Change...
InstallShield
< Back
Next >
Cancel

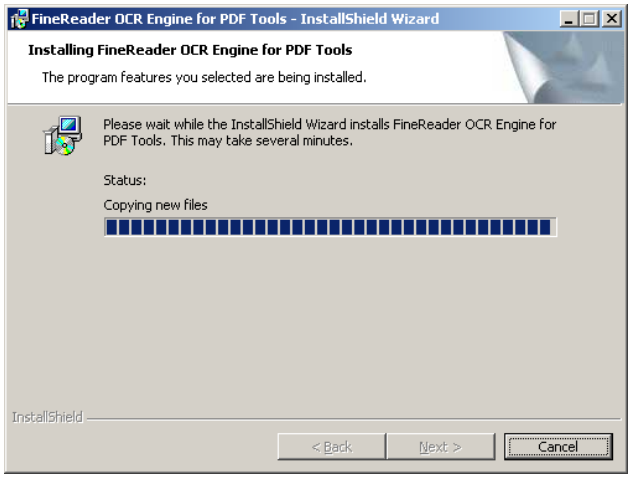
6. Enter the Serial Number. The serial number is provided by PDF Tools AG. The serial number is required for any license type (evaluation license, page limited license, unlimited license). If you do not have a serial number at this point, you can continue without entering the serial number. You can provide the serial number at a later time. Press "Next" when done.


FineReader OCR Engine for PDF Tools - InstallShield Wizard
Serial Number
Enter the serial number for the ABBYY FineReader OCR Engine
Serial Number:
FECN-
You can enter the serial number for the ABBYY FineReader OCR Engine now. You can still enter this information later when using the License Manager utility to activate the license for this installation.
InstallShield
< Back
Next >
Cancel

7. When ready to install, press the “Install” button in the next dialog.


FineReader OCR Engine for PDF Tools - InstallShield Wizard
Ready to Install the Program
The wizard is ready to begin installation.
Click Install to begin the installation.
If you want to review or change any of your installation settings, click Back. Click Cancel to exit the wizard.
InstallShield
< Back
Install
Cancel

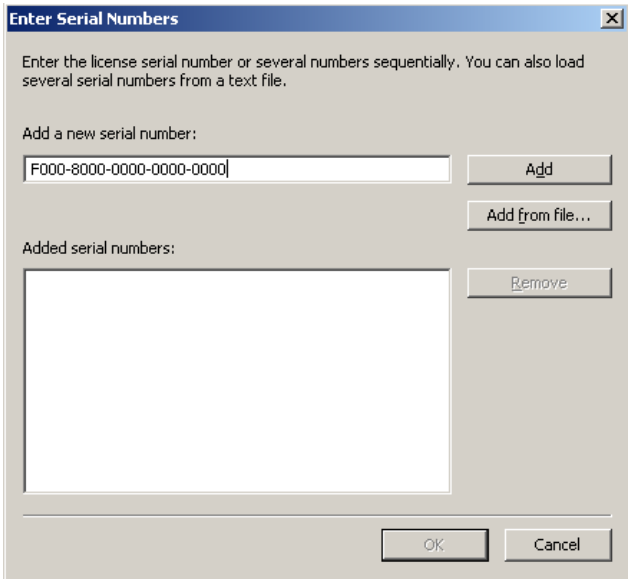
8. The Fine Reader OCR Engine for PDF Tools is now being installed.


FineReader OCR Engine for PDF Tools - InstallShield Wizard
Installing FineReader OCR Engine for PDF Tools
The program features you selected are being installed.
Please wait while the InstallShield Wizard installs FineReader OCR Engine for PDF Tools. This may take several minutes.
Status:
Copying new files
InstallShield
< Back
Next >
Cancel

License Activation of ABBYY OCR FineReader Engine

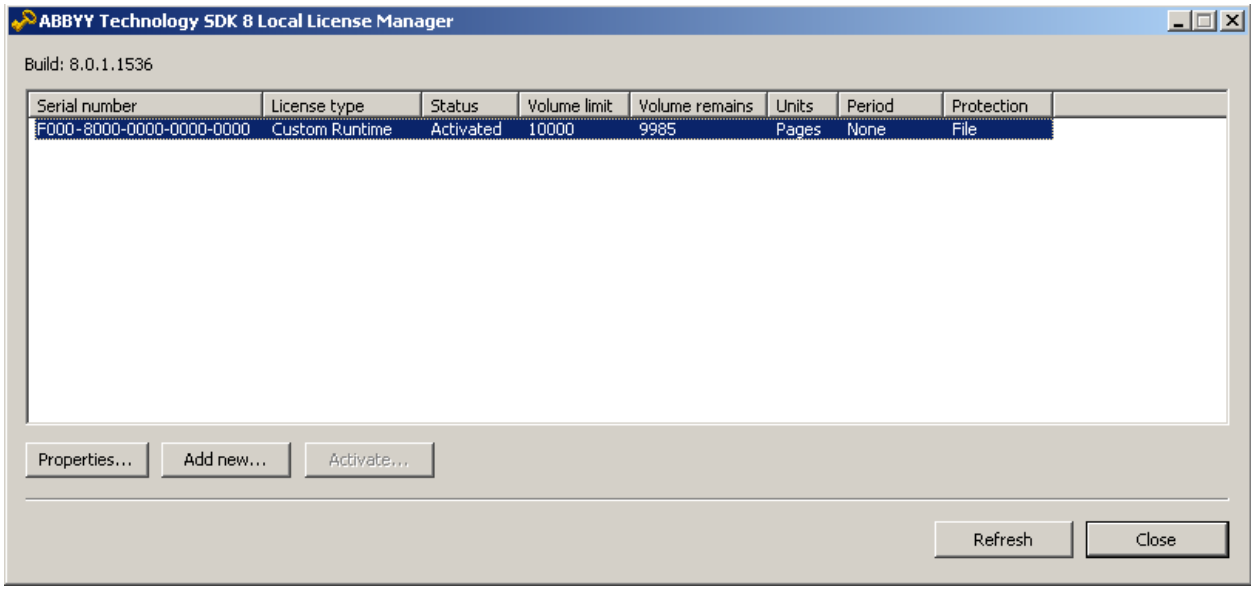
After the installation, the license needs to be activated. This is done by doing the following steps.

1. Launch the License Manager. Go to Start ->Programs ->PDF Tools AG ->FineReader OCR Engine for PDF Tools AG ->Launch License Manager



   If you haven't provided the serial number at the time of installation, a dialog window asking for the serial number is shown. The serial number has a format like this:
   F000-8000-0000-0000-0000

2. If a valid serial number is provided, the next dialog is shown. Press the button "Add..." in order to activate the serial number. Note that every serial number (including evaluation versions) can only be activated once.



   Should you accidentally have provided a wrong, but not necessarily an invalid serial number or want to change the serial number for another reason, do the following steps:

   - Go to the directory "*C:\Program Files\PDF Tools AG\FREngine\Bin*" and delete the file *FREngine.ini*
   - If in the same directory there is a file with a file name equal to your old license key, something similar to "F000-8000-0000-0000-0000.lprf", delete that file.
   - Launch the License Manager.
   - Enter the new serial number in the dialog window.

Upgrading to a Newer Version

If a newer version of the Abbyy FineReader Engine is available and you would like to upgrade to that, do the following steps:

- Even though these files should be left alone, it's safer to make a backup of your configuration file and your license key file. These two files can be found in the directory "*C:\Program Files\PDF Tools AG\ FREngine\Bin*" and are called something similar to "F000-8000-0000-0000-0000.lprf" and "*FREngine.ini*".
- Uninstall the FR Engine using the appropriate command from the Start menu
- Install the new version of the FR Engine

Remove a Test License

If you want to replace a previously installed test license with the production license, you need to deactivate or delete the test license. If deactivation via the license manager tool fails, follow the following steps.

Abbyy FineReader 8 licenses can be removed by deleting the corresponding license file in the installation's Bin folder.

Abbyy FineReader 10 and 11 license information is stored in the *%AllUsersProfile%\Application data\ABBYY\SDK\10\ Licenses* resp. *%AllUsersProfile%\Application data\ABBYY\SDK\11\ Licenses* folder. The license file is named according to the serial key; remove this file. After this, you need to restart the licensing service to expedite detection of the configuration change.

2.2 OCR Service

If you are using the 3-Heights™ OCR Enterprise Add-On in combination with a service, such as:

- 3-Heights™ Image to PDF Converter Service
- 3-Heights™ PDF to PDF/A Converter Service
- 3-Heights™ Document Converter

the OCR engine must be invoked via the 3-Heights™ OCR Service and not directly by the other service.

Installation of the OCR Service

The 3-Heights™ OCR Service can be downloaded from your customer account at www.pdf-tools.com.

1. Download the ZIP archive "OCRSRV<version><platform>.zip" (e.g. OCRSRV450WIN32.zip) from your download account
2. Extract the file 3-Heights(TM) OCR Service.msi from the ZIP archive.
3. Double-click the MSI file to start the installation wizard.
4. Follow the installation wizard.

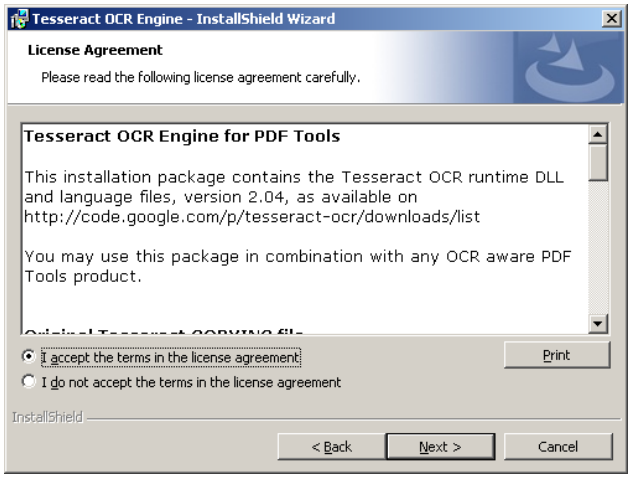
Please refer to the manual of the OCR service (*OcrSvr.pdf*) for more information.

2.3 Tesseract for Windows

1. Download the ZIP archive using the following link: http://www.pdf-tools.com/public/downloads/resources/OCRTESSWIN32-64.zip
2. Extract the file *Tesseract OCR Engine.msi* from the ZIP archive.
3. Double-click the MSI file to start the installation wizard.
4. In the first dialog box of the installation wizard press the button “Next”.

Tesseract OCR Engine - InstallShield Wizard
Welcome to the InstallShield Wizard for Tesseract OCR Engine
The InstallShield(R) Wizard will install Tesseract OCR Engine on your computer. To continue, click Next.
WARNING: This program is protected by copyright law and international treaties.
< Back
Next >
Cancel

5. Follow the wizard.


Tesseract OCR Engine - InstallShield Wizard
License Agreement
Please read the following license agreement carefully.
Tesseract OCR Engine for PDF Tools
This installation package contains the Tesseract OCR runtime DLL and language files, version 2.04, as available on http://code.google.com/p/tesseract-ocr/downloads/list
You may use this package in combination with any OCR aware PDF Tools product.
I accept the terms in the license agreement
I do not accept the terms in the license agreement
Print
InstallShield
< Back
Next >
Cancel

The installation process defines new environment variables, which means the process (application or shell) that does OCR needs to be restarted.

2.4 Tesseract for Unix

Introduction

OCR aware PDF Tools products have the ability to dynamically load the Tesseract OCR plug-in on supported Unix platforms. To make use of the OCR feature, it is mandatory to also install the Tesseract language support files.

At this time Tesseract is supported for Linux (kernel V 2.6 builds).

Installation

Download the Tesseract engine support files from the PDF Tools’ website https://www.pdf-tools.com/pdf/resources.aspx

- Tesseract301-tessdata.tar.gz (Tesseract 3 data)
- Tesseract301-linux-i386.tar.gz (Linux 32-bit shared libraries)
- Tesseract301-linux-x64.tar.gz (Linux 64-bit shared libraries)

Extract the files in the archive to a suitable location. During extraction, a 'tessdata' folder will be created. A convenient place for storing this tessdata folder is /usr/local/share as this is the default location where the engine will look for the data.

Note that the Tesseract library libtesseract.so depends on liblept.so which again depends on libz.so and libpng12.so. Our download packages contain liblept.so, but not the libz.so and libpng12.so, as these are pre-installed on most Linux platforms. Please make sure that all required standard libraries are installed by using the 'ldd' command on libtesseract.so.

Runtime Configuration

Unless having the 'tessdata' folder located in one of the directories listed in the PATH environment variable, you need to define the environment variable TESSDATA_PREFIX to point to the directory containing the 'tessdata' folder. Make sure the TESSDATA_PREFIX definition ends with a slash, as the Tesseract engine will directly concatenate its value with "tessdata".

2.5 Installing Products Using the Add-On

The 3-Heights™ OCR Enterprise Add-On can be used in combination with one of the following products. Other applications cannot access of the 3-Heights™ OCR Enterprise Add-On:

- 3-Heights™ Image to PDF Converter (Shell, Service, API)
- 3-Heights™ PDF to PDF/A Converter (Shell, Service, API)
- 3-Heights™ Document Converter

How these products are installed is described in the corresponding manuals of these products.

Linux

Linux product kits do not contain any OCR plugins. However, selected plugins are available separately for selected UNIX platforms including major Linux versions. As of now, PDF Tools provides plugins for Tesseract 3.0 and for the 3-Heights™ OCR Service.

The Tesseract plugin is named "libpdfocrpluginTesseract.ocr". The OCR Service plugin is named "libpdfocrpluginService.ocr".

Unlike on Windows, you do not need to install OCR plugins into the same directory as the tool using it. Plugins are rather located via the PDFOCRPLUGINPATH environment variable. When this environment variable is not set, it defaults to ${HOME}/.pdf-tools/ocr-plugins:/etc/opt/pdf-tools/ocr-plugin

Also make sure to properly configure the PDFFONTDIR environment variable to point to a folder containing a collection of font files that include serif, non-serif and monospaced fonts. Note: whenever there is a change in the configuration of PDFFONTDIR (including fonts added or removed in one of the directories listed), make sure to delete the font-database*dat files in the tmp folder (folder defined by PDFTMPDIR environment variable, TMP environment variable, or /tmp).

When using the OCR Service plugin, also make sure that the ocrserver.ini configuration file is properly configured and also stored in one of the directories listed in PDFOCRPLUGINPATH. (Only HTTP-based service entries are supported).

3 Licensing

3.1 ABBYY FineReader Engine

This is an overview of the licenses available. Details can be retrieved from a sales representative of PDF Tools AG at pdfsales@pdf-tools.com.

License Types

The type of license is defined in the serial key. There are 3 runtime license types available:

Runtime License Options	Network Options	Features

▪ Renewable Volume Licenses
  ▪ Volume per time period
▪ CPU Core Licenses
  ▪ No volume restriction
▪ Project-Based Licenses
  ▪ Volume per project, unlimited scalability
	▪ 1 Machine Licenses
	▪ Network Licences.
	Allow distributing the licensed:
	  ▪ Number of cores
	  ▪ Page volumes
		▪ Standard Features
		▪ Standard & Add Ons Features, e.g.
		  ▪ OCR Languages
		  ▪ ICR (handprint recoginition)

Renewable Volume Runtime Licenses

Volume pricing is based on the quantity of pages that can be processed in a certain period. The renewable period can be a month or a year. Volume licenses have no CPU core limitation. This network licensing allows the dynamic sharing of the licensed volume between multiple network computers, network license manager required.

CPU Core Runtime Licenses

CPU core pricing is based on the quantity of CPU cores that can be used for processing. These license have no page limitation. CPU core licenses are available for 1 machine or the number of licensed cores can be distributed over the network.

Project-Based Licenses

The project pricing scheme allows processing of a defined volume. Efficiency is guaranteed, because of unlimited number of multi-core machines can be use.

4 Reference Manual

4.1 Architecture

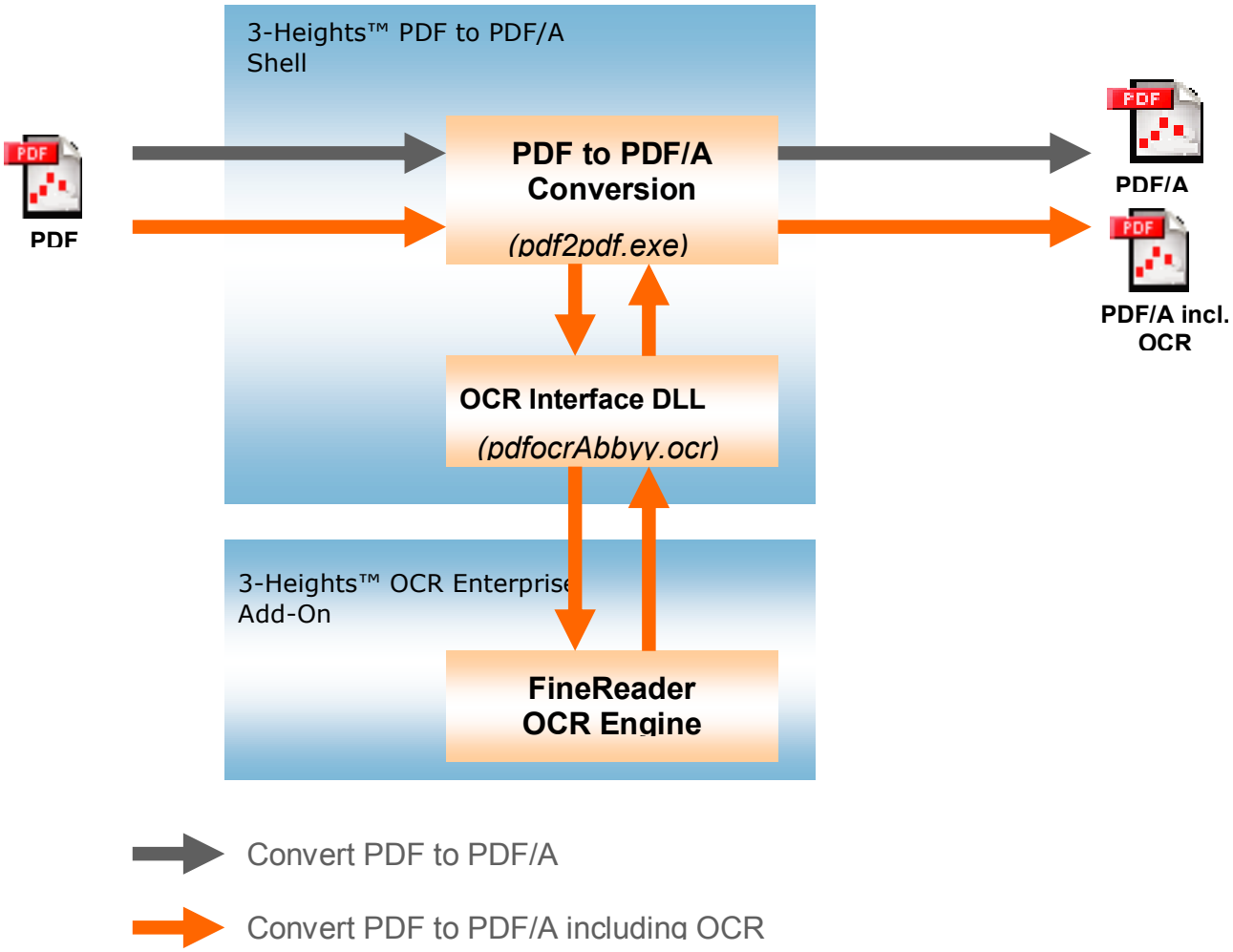
Use OCR Engine Directly

How the 3-Heights™ OCR Enterprise Add-On is used in combination with another product of the 3-Heights™ line is shown in this chapter using the 3-Heights™ PDF to PDF/A Shell as an example. The product can be used in the same way with the 3-Heights™ Image to PDF Converter Shell/API and the 3-Heights™ PDF to PDF/A Converter API.

The 3-Heights™ OCR Enterprise Add-On installs the OCR engine which can be invoked from the 3-Heights™ PDF to PDF/A Shell.

The OCR interface DLL (e.g. pdfocrAbbyy.ocr) that communicates with the OCR engine is distributed with the 3-Heights™ PDF to PDF/A Shell. The OCR interface DLL does not contain any OCR technology itself.

It is an interface between the PDF to PDF/A Shell and the 3-Heights™ OCR Enterprise Add-On.

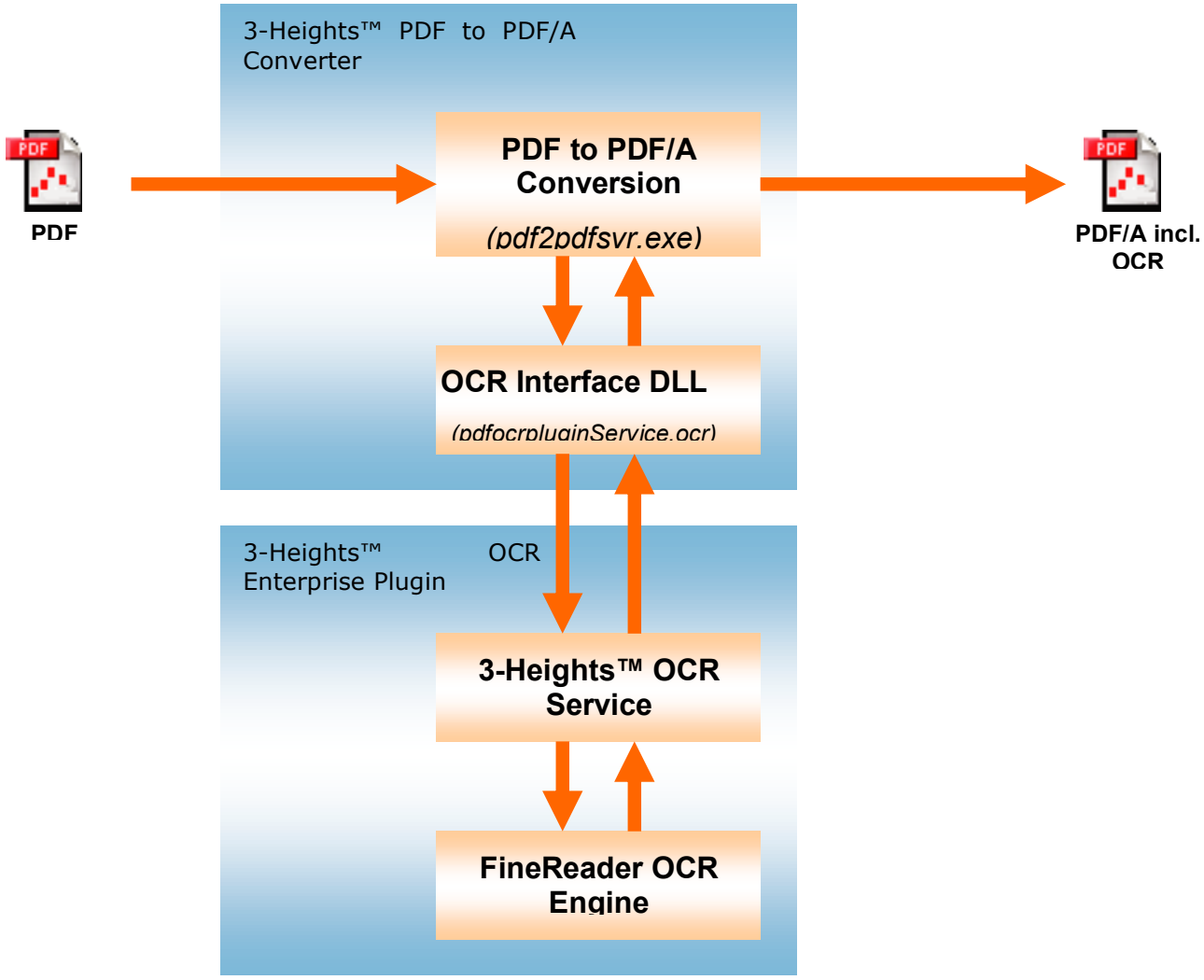


Once the 3-Heights™ OCR Enterprise Add-On is installed, it can be used by other applications of the 3-Heights™ line. How these other application make use of the 3-Heights™ OCR Enterprise Add-On is individual for every product.

Use OCR Engine via Service

Instead of the application invoking the OCR engine directly, the OCR engine can be accessed via a service. The interface DLL that communicates with the OCR service is *pdfocrpluginService.ocr*. The service invokes the OCR engine and returns the result back to the application. Using the service has the following advantages:

- If the application processes files in parallel, the requests to the OCR engine need to be serialized to match its technical capabilities and licensing limitations.
- The application and the OCR engine can reside on different computers to improve load-balancing.



All 3-Heights™ services must use the OCR service and not the OCR engine directly. Shell and API products can use either the OCR service or the OCR engine directly.

4.2 OCR Related Switches and Functions

The 3-Heights™ OCR Enterprise Add-On can be used with the 3-Heights™ PDF to PDF/A Converter and the 3-Heights™ Image to PDF Converter. Both tools are available as shell, Windows service and as API version. The switches of shell and service version and the functions of the API version that are related to OCR are described in Table: Switches and Functions.

Table: Switches and Functions

Shell/Service	API	Description
`-le`	`GetOCREngineCount()` `GetOCREngine(int)`	List available OCR engines. OCR engines are listed by means of the interface DLLs. The OCR Interface DLLs are distributed with converter and are named *.ocr, e. g. *pdfocrAbbyy.ocr*. The shell version lists all available OCR Interface DLLs with the switch -le. In the API, the number of OCR engines is retrieved using GetOCREngineCount. The string of the nth engine is retrieved using GetOCREngine(n). The name of the Abbyy FineReader OCR Engine 8.1 is "abbyy". The name of the 3-Heights™ OCR Service is "service". The name of the Tesseract OCR Engine is "tesseract". This switch it not available for Windows services products, because they only support one engine type: "service".
`-ocr name`	`SetOCREngine(String)`	Set OCR engine. The name of the engine is the string previously retrieved using the listing function.
`-ocl lang`	`SetOCRLanguage(String)`	Set OCR language(s). This setting is OCR engine dependant.
`-ocp params`	`SetOCRParams(String)`	Set OCR engine dependant parameters. A parameter is always a key-value pair. Multiple pairs can be set as one comma-separated string.

4.3 Examples

This chapter describes how the 3-Heights™ OCR Enterprise Add-On is accessed from another product.

The shell samples in this chapter use the 3-Heights™ PDF to PDF/A Converter Shell (pdf2pdf). The commands for the 3-Heights™ Image to PDF Shell (img2pdf) are equivalent with respect to OCR. That is also true for the service version of both products, with the difference that the settings are not written to the console, but are applied in the configuration file instead.

Set OCR Engine

3-Heights™ products that support OCR allow for listing and selecting available OCR engines. At this time, the OCR engine "Abbyy FineReader 8.1", "Abbyy FineReader 10", the OCR service and tesseract are supported.

Shell

The following shell sample shows how to convert a PDF to a PDF/A document without applying OCR.

Example: Converting a PDF document to PDF/A without OCR does not require the 3-Heights™ OCR Enterprise Add-On.

```
pdf2pdf -cl pdfa-1b input.pdf output.pdf
```

OCR engines are listed using the switch -le. In order to list an OCR engine, its OCR interface DLL must be in the same directory as where the executable reside. OCR interface-DLLs have the file extension ".ocr".

Example: The listing function does not detect whether an OCR engine is actually installed.

```
pdf2pdf -le
List of available OCR engines:
- abbyy
- abbyy10
- service
- tesseract
End of list.
```

Once the name of the engine is known, it can be provided as parameter to the -ocr switch for converting.

Example: Converting a PDF document to a PDF/A document and add OCR information requires the 3-Heights™ OCR Enterprise Add-On to be installed.

```
pdf2pdf -cl pdfa-1b -ocr "abbyy" input.pdf output.pdf
```

Service

In a service product (e.g. 3-Heights™ PDF to PDF/A Service), the OCR service must be defined as "service".

Example: Converting a PDF document to PDF/A document and add OCR information using the 3-Heights™ PDF to PDF/A Converter Service requires the 3-Heights™ OCR Enterprise Add-On including the OCR service to be installed. The OCR engine is defined in the configuration file of the main application ("*pdf2pdfsvr.ini*").

```
-cl pdfa-1b -ocr "service"
```

API

In the API version, there are three relevant calls, which are used in the following sequence:

- `nCount = GetOCREngineCount` to retrieve the number of available OCR engines
- `Name = GetOCREngine(n)` to retrieve the name of the nth OCR engine, where n is a value from 1 to nCount.
- `SetOCREngine(Name)` to set the OCR engine

If the name of the OCR engine is known, the first two calls can be omitted.

Example: Set OCR engine if its name is known.

```
SetOCREngine("abbyy")
```

Set OCR Language

Multiple languages can be set as one comma-separated string. The supported languages are listed in the appendix of this manual. Setting a language helps the OCR Engine to minimize errors by means of using dictionaries of the defined languages. Setting the language also improves the speed. Abbyy recommends to always set one or several languages, but not more than six. The default language is English.

Example: Shell (Abbyy) - Set the languages to English and German.

```
pdf2pdf -ocr "abbyy" -ocl "English, German" input.pdf output.pdf
```

Example: Shell (Tesseract) - Set the language to German.

```
pdf2pdf -ocr "tesseract" -ocl "deu" input.pdf output.pdf
```

Example: API (Abbyy) - Set the languages to English and German.

```
SetOCREngine("abbyy")
SetOCRLanguage("English, German")
```

This setting is optional.

Set OCR Engine Specific Parameters

Set OCR engine specific parameters. The parameter is a series of key-value pairs. The syntax of a key-value pair is: “key1=value1”. Multiple pairs can be set as one comma-separated string: “key1=value1, key2=value2”. The supported parameters are listed in the appendix of this manual.

Example: (Shell) Set the cell count to 2 (this is used for tables) and enable the balanced mode.

```
pdf2pdf -ocr "abbyy10" -ocp "Profile=C:\ocr.ini" input.pdf output.pdf
```

Example: (API)

```
SetOCREngine("abbyy")
SetOCRParams("CellCount=2, BalancedMode=TRUE")
```

This setting is optional. This setting re-initializes the OCR languages to default. It must therefore be set before the OCR language.

4.4 Impacts of Using OCR

Applying OCR to a document compared to not applying OCR has mainly the following consequences.

Performance

OCR takes roughly 2-3 seconds per A4 page on a decent CPU. That is much more than the conversion of the content from one format to another (e.g. from PDF to PDF/A, or from an image to PDF/A). Therefore, in most cases, the OCR-process is the bottleneck for the throughput.

Performance can be improved by using multiple CPUs or configuring the OCR-process to run less accurate and thereby faster (e.g. FastMode, or BalancedMode).

See also chapter Performance Figures.

File Size

Depending on the amount of text found, the file size is increased by roughly 0-10 KB for an A4 page.

Appearance

In general, when applying OCR, the original image in the document is replaced by the processed image that was used by the OCR-engine for text recognition. This image is de-skewed and possibly de-noised or has been modified in other ways.

As a consequence the text fits perfectly. However, the image may have visual changes.

If a main attribute of the image (color space, resolution) was changed by the OCR-engine for the purpose of text recognition, the original image is kept. In this case, it is possible that some text is not matching exactly (translation/skewing by a few pixels).

5 Configuration and Performance

5.1 Configuration for ABBYY

The Abbyy FineReader OCR Engine 8.1 is optimized for A4 or smaller sized pages. Applying OCR to one A4 page, which has a resolution of 300 dpi, takes about 2-3 seconds. Larger pages or pages with higher resolution require more memory and time to apply OCR. Abbyy makes the following recommendations:

- Images should have a resolution of around 300 dpi
- Images with small text (font size under 10 points) should have a resolution of 400 to 600 dpi
- Grayscale images are suited best of OCR
- Setting the correct language improves the recognition rate and speed; Recommended is setting 3 to 6 languages
- Using the FastMode improve the throughput at the cost of recognition rate
- Using the normal mode yields the best recognition rate
- The BalancedMode is between the FastMode and the normal mode

5.2 Configuration for Tesseract

Tesseract is optimized for images with a resolution of 300 dpi. Any image with a different resolution will take extra time.

5.3 Performance Overview

In the process of converting images to PDF/A, or PDF to PDF/A including OCR information there are basically three figures that are of interest. Depending on the process, one of these figures may be more important than others: Throughput, recognition rate and file size

Throughput

How long does it take to convert and apply OCR to one page? This figure is particularly important when dealing with large volumes, such as converting entire TIFF archives to PDF/A. Normally the throughput is measured in pages per minute (ppm). How long it takes to process a page depends on various factors. The most important are:

1. CPU Power: A faster CPU processes the same file faster than a slower CPU.
2. Number of CPUs: Whilst the 3-Heights™ OCR Enterprise Add-On does not support dividing one job to multiple CPUs, it does support processing multiple files in parallel. Thus if two or more CPUs are available, the load can be spread.
3. Input file: Generally a larger image requires more processing time than a smaller image. The size of the image is given by the number of pixels.
4. Settings of OCR Engine: An OCR engine can have different modes, which allow the user to choose a tradeoff between throughput and recognition rate. Normally it also helps to set the correct language of the text to which OCR is to be applied.

Recognition State

How much text can be identified depends on different factors. The most important one is the document itself. A low resolution document with text that is hardly readable by eye will result in a much lower recognition rate than a clear scan with non-skewed black text on white background.

According to the Abbyy documentation its OCR engine has a rate of 1-2 errors per page.

File Size

How large is the resulting file compared to the original image file?

Applying OCR to a document which contains scanned pages increases the file size by around 10%. At the same time it is possible to change the compression algorithm of the images and thereby reduce the file size. Changing the compression from G4 (which is supported in TIFF, PDF and PDF/A) to JBIG2 (which is supported by PDF and PDF/A) reduces the size of the images by around 50%.

5.4 Performance Figures

In this chapter we run some test the ABBYY FineReader 8 (build 8.0.1.1643) and 10 (build 10.0.3.494) and Tesseract engines with different types of documents to find out some real figures. We focus on the throughput in pages per minute (ppm).

Test Scenario

Input Documents:
We use two different types of input documents:

1. *Multi-page TIFF files*, which contain the first 100 pages of the PDF Reference Manual 1.4. The TIFF files are created electronically using a high-end rendering engine that applies anti-aliasing and optionally dithering. As a result these test files are of very high visual quality, compared to scanned images. We use two different resolutions (150 dpi and 300 dpi) and two different compression types (G4 and JPEG).
   For the test these TIFF files are processed by the 3-Heights™ Image to PDF Converter using the OCR Add-On.
2. A PDF test set, which contain scans of English text. Each PDF file consists of three to eight pages. Each page contains a scanned image, of which some are skewed. All images have a resolution between 150 and 300 dpi. Half the images are grayscale, the others are color.
   For the test these PDF files are batch converted to PDF/A using the 3-Heights™ PDF to PDF/A Converter with the OCR Add-On.

Test Systems:
The tests are run on two different systems:
System 1: Core i7, 2.93 GHz, 4 GB, Windows 7 (Performance Information: CPU 7.5, RAM 7.5, HD 5.9)
System 2: Pentium 4, 3.0 GHz, 512 MB, Windows 2000 Server In this test we only ever use one thread and therefore only one CPU core.
Throughout all tests system 2 was constantly 3 - 3.1 times slower than system 1. Therefore only the numbers of system 1 are published in this chapter.

No OCR

Converting TIFF to PDF/A, or PDF to PDF/A without applying OCR goes rather quickly. On average we are looking at a throughput of around 5000 pages per minute (ppm). There are scenarios where the conversion is even faster: If images do not need re-encoding (e.g. JPEG), the throughput goes up to 100'000 ppm. On the other hand there are scenarios where the conversion takes longer: If changing the compression of bi-tonal images from G4 to JBIG2, the throughput is lowered to around 800 ppm.

All in all these numbers are not that important due to the fact the OCR requires considerably more time than the format conversion, as we will see in the next chapter.

ABBYY FineReader 8

TIFF files at 150 dpi: 53-63 ppm
TIFF files at 300 dpi: 100 ppm[1]
PDF test set (150-300 dpi, skewed): 52 ppm

- The throughput is similar for color and grayscale images
- The throughput is similar for dithered and non-dithered image
- Not setting the correct language: 1.5 times slower, recognition rate is noticeably lower
- Using FastMode: 1.6 - 2.3 times faster (the better the quality of the input file, the smaller the difference between normal and fast mode for both, speed and throughput)
- Using BalancedMode: 1.3 - 2.0 times faster
- The file size increases by around 4 KB per page
- The recognition rate is quite steady and usually at 0-3 errors per page. However on some pages there are fragments, such as a title or a table, which seem to be ignored.

ABBYY FineReader 10

TIFF files at 150 dpi: 32-36 ppm
TIFF files at 300 dpi: 41 ppm
PDF test set (150-300 dpi, skewed): 37 ppm

- The throughput is similar for color and grayscale images
- The throughput for dithered images is about 20% faster
- Not setting the correct language has no impact
- Using Predefined Profiles: 1.1 - 1.3 times faster than no default settings
- The file size increases by around 3 KB per page
- The recognition rate is quite similar to ABBYY FineReader Engine 8.

Tesseract

TIFF files at 150 dpi: 17 - 21 ppm
TIFF files at 300 dpi: 24 - 29[2] ppm
PDF test set (200 dpi, skewed): 33 ppm

- G4 compressed images take about 1.2 times longer than JPEG compressed images
- Dithered images take around 1.2 times longer than non-dithered images
- Not setting the correct language: 1.1 times slower
- The file size increases by around 7 KB per page
- The recognition rate varies heavily. For images at 300 dpi with a good visual quality there are around 3 errors per page. For lower resolutions or more “difficult” images the error rate increases to 10-20 errors per page.

Summary

For scans of high visual quality and a resolution of 300 dpi the ABBYY and Tesseract engine deliver similar results. For any other scenario the ABBYY engine turns out to be faster and have a better recognition rate.

Tests have shown that the throughput is directly related to the CPU power. ABBYY generally requires more RAM than Tesseract, however in the tests the total amount of memory remained between 20 and 80 MB at all times.

Approximately how long does it take to convert 1 million A4 documents on a dedicated system?

[1] Older builds (prior to 8.0.1.1643) of the ABBYY FR Engine 8 were considerable slower at 300 dpi.
[2] Tested showed that Tesseract runs faster at 300 dpi than at 150 dpi

OCR Engine	P4, 3.0 Ghz, 1CPU	i7, 2.93 Ghz, 1CPU	i7, 2.93 Ghz, 2CPU
150dpi			
ABBYY FineReader 8	37 days	12 days	6 days
ABBYY FineReader 10	64 days	21 days	10 days
Tesseract	125 days	41 days	21 days
300dpi			
ABBYY FineReader 8	22 days	7 days	4 days
ABBYY FineReader 10	53 days	17 days	9 days
Tesseract	89 days	29 days	15 days

All values presented here are average values of a series of test runs. Other scenarios with other documents may yield different results.

6 Licensing and Copyright

The 3-Heights™ OCR Enterprise Add-On is copyrighted. This user's manual is also copyright protected; it may be copied and given away provided that it remains unchanged including the copyright notice.

7 Contact

PDF Tools AG

Kasernenstrasse 1

8184 Bachenbülach

Switzerland

http://www.pdf-tools.com

A ABBYY FineReader 8

Note: Information in this appendix base on the "ABBYY FineReader User Guide". The syntax of all parameters and certain default values however are different.

[3] 2 CPUs are used when starting two individual threads

A.1 Languages

Please note:

- Not all languages are supported in the base version. See also chapter "Licensing".
- The Engine does not support language combinations consisting of two hieroglyphic languages or combinations of a hieroglyphic language and a non-hieroglyphic language. However, the Engine will recognize bilingual texts written in English and a hieroglyphic language - when you select "ChineseTraditional", "ChineseSimplified", "Japanese" or "Korean", the Engine will add the "English" recognition language automatically.

The following languages are supported:

Abkhaz
Adyghe
Afrikaans
Agul
Albanian
Altaic
ArmenianEastern
ArmenianGrabar
ArmenianWestern
Awar
Aymara
AzeriCyrillic
AzeriLatin
Bashkir
Basque
Belarusian
Bemba
Blackfoot
Breton
Bugotu
Bulgarian
Buryat
Catalan
Chamorro
Chechen
ChineseSimplified
ChineseTraditional
Chukcha
Chuvash
Corsican
CrimeanTatar
Croatian
Crow
Czech
Danish
Dargwa
Dungan
Dutch
DutchBelgian
English
EskimoCyrillic
EskimoLatin
Esperanto
Estonian
Even
Evenki
Faeroese
Fijian
Finnish
French
Frisian
Friulian
GaelicScottish
Gagauz
Galician
Ganda
German
GermanNewSpelling
GermanLuxembourg
Greek
Guarani
Hani
Hausa
Hawaiian
Hebrew
Hungarian
Icelandic
Ido
Indonesian
Ingush
Interlingua
Irish
Italian
Japanese
Kabardian
Kalmyk
KarachayBalkar
Karakalpak
Kasub
Kawa
Kazakh
Khakas
Khanty
Kikuyu
Kirgiz
Kongo
Korean
Koryak
Kpelle
Kumyk
Kurdish
Lak
Lappish
Latin
Latvian
Lezgin
Lithuanian
Luba
Macedonian
Malagasy
Malay
Malinke
Maltese
Mansi
Maori
Mari
Maya
Miao
Minankabaw
Mohawk
Moldavian
Mongol
Mordvin
Nahuatl
Nenets
Nivkh
Nogay
Norwegian
NorwegianBokmal
NorwegianNynorsk
Nyanja
Occidental
Ojibway
OldEnglish
OldFrench
OldGerman
OldItalian
OldSpanish
Ossetic
Papiamento
PidginEnglish
Polish
PortugueseBrazilian
PortugueseStandard
Provencal
Quechua
RhaetoRomanic
Romanian
RomanianMoldavia
Romany
Ruanda
Rundi
RussianOldSpelling
Russian
Samoan
Selkup
SerbianCyrillic
SerbianLatin
Shona
Sioux
Slovak
Slovenian
Somali
Sorbian
Sotho
Spanish
Sunda
Swahili
Swazi
Swedish
Tabassaran
Tagalog
Tahitian
Tajik
Tatar
Thai
Tinpo
Tongan
Tswana
Tun
Turkish
Turkmen

Tuvin	Visayan	Basic	E13B
Udmurt	Welsh	C++	CMC7
UighurCyrillic	Wolof	Cobol	Mixed
UighurLatin	Xhosa	Fortran	Digits
Ukrainian	Yakut	Java	
UzbekCyrillic	Zapotec	Pascal	
UzbekLatin	Zulu	Chemistry	

A.2 OCR Engine Parameters

This chapter lists all supported key-value pairs for the ABBYY FR Engine 8. Some keys are also valid for the FR 10. Values written in bold indicate the default value of this key. Neither the keys nor the values are case-sensitive. Certain keys allow for an OR superposition. OR super-positions are applied by repeating the key.

Example: Set the orientation to Left_To_Right OR Down_To_Top.

```
-ocp "Orientation=Left_To_Right, Orientation=Down_To_Top"
```

Barcode Parameters

Table: Barcode Parameters

Key	Value	Description
HasChecksum	TRUE, **FALSE**	Specifies whether the barcode being recognized must be interpreted as the barcode of the same type but with a check sum. Only available for barcodes of types Code 39, Interleaved 2 of 5, and Codabar.
IsCode39WithoutAsterisk	TRUE, **FALSE**	Specifies that the Code 39 barcode being recognized has no start and stop symbol, the asterisk "*". Only available for barcodes of the Code 39 type.
IsEAN13InterpretedAsUPCA	TRUE, **FALSE**	Specifies whether the EAN 13 barcode being recognized should be interpreted as a barcode of type UPC A. Only available for barcodes of type EAN 13. Note: UPC A barcodes are EAN 13 barcodes with the <0> first character.
Orientation	**Autodetect**, Unknown, Left_To_Right, Down_To_Top, Right_To_Left, Top_To_Down	Specifies an OR superposition of the constants which denote the types of barcode orientation. For example, if it is set to Left_To_Right OR Down_To_Top, ABBYY FineReader Engine will presume that barcode blocks may be oriented either from left to right or from down to top, ignoring all other variants. The default is set to Autodetect, i.e. ABBYY FineReader Engine will detect the barcode orientation automatically.

PDF417CodePage	**Null**, Latin, Cyrillic, EasternEuropean, Baltic, Turkish, US_MSDOS, LatinI_MSDOS, Russian_MSDOS, Baltic_MSDOS, Turkish_IBM, Slavic_MSDOS, Greek, Greek_737, Greek_869, Latin_ISO,EasternEur Turkish_ISO, Baltic_ISO, Cyrillic_ISO, Greek_ISO, KOI8, Tatar, Tatar_MSDOS, Roman_Macintosh, Greek_Macintosh, Cyrillic_Macintosh, Ukrainian_Macintosh Latin2_Macintosh, Icelandic_Macintosh, Turkish_Macintosh, Croatian_Macintosh, Armenian, Armenian_MSDOS, Armenian_Macintosh, Latin5_ISO, Cyrillic_MSDOS, Bashkir	Used to recognize barcodes which do not conform to the barcode specifications. Do not use for barcodes created in conformity with the barcode specifications. Some barcode printers use code pages other than US_MSDOS required by the specifications. In this case, use this setting to specify the code page which was used by the barcode printer to create the barcode. In most cases this will be the code page of the operating system under which the barcode printer was running.
SupplementType	**Autodetect**, Unknown, Void, 2Digits, 5Digits	Specifies an OR superposition of the constants which denote the supplement type. This property is only available for barcodes of the EAN 8, 13 and UPC E types. For example, if it is set to Void OR 2Digits, ABBYY FineReader Engine will try to recognize barcode blocks either without supplementary barcode or with 2 digit supplementary barcode. The default is set to Autodetect, i.e. ABBYY FineReader Engine will detect the supplementary barcode type automatically.
Type	**Autodetect**, Unknown, Code39, Interleaved25, EAN13, Code128, EAN8, PDF417, Codabar, UPCE, Industrial25, IATA25, Matrix25, Code93, PostNet, UCC128	Specifies an OR superposition of the constants which denote the types of barcode. For example, if it is set to EAN13 OR EAN8, ABBYY FineReader Engine will try to recognize barcode blocks in either EAN 13 or EAN 8 standard, ignoring all other variants. The default is set to Autodetect, i.e. ABBYY FineReader Engine will detect the barcode type automatically.

Image Processing Parameters

Table: Image Processing Parameters		
Key	**Value**	**Description**
AutodetectInversion	TRUE, **FALSE**	Specifies whether the color inversion must be automatically detected during preprocessing. If colors were inverted the InvertImage key will be set to TRUE. If this keyis TRUE the InvertImage key is ignored.
BlackGarbageSize	**0**	Specifies the size in pixels of black dots which must be considered garbage. The default value of this 0, which means that the size will be set automatically. If you change the value to a positive number, the RemoveGarbage key will automatically become TRUE.
InvertImage	TRUE, **FALSE**	Specifies if the image colors are inverted during preprocessing. The value of this key is taken into account only if AutodetectInversion key is FALSE.
MirrorImage	TRUE, **FALSE**	Specifies if the image is mirrored around the vertical axis during preprocessing.
ProhibitCorrectLocalSkew	TRUE, **FALSE**	Correction of local skew is performed in case the skew for the whole page is calculated uncertainly. This key allows you to prohibit local skew correction.
RemoveGarbage	TRUE, **FALSE**	Specifies if garbage (excess dots that are smaller than a certain size) is to be removed from the image during preprocessing.
RotationType	**NoRotation**, Clockwise, Counterclockwise, Upsidedown	Specifies what type of rotation will be performed upon the image during its preprocessing. This key is NoRotation by default, which means that image is not rotated.
WhiteGarbageSize	**0**	Specifies the size in pixels of white dots which must be considered garbage. The default value of this 0, which means that the size will be set automatically. If you change the value to a positive number, the RemoveGarbage key will automatically become TRUE.

Page Analysis Parameters

Table: Page Analysis Parameters		
Key	**Value**	**Description**
DetectBarcodes	TRUE, **FALSE**	Specifies if barcodes are detected, and accordingly barcode blocks created, during the page layout analysis. If this property is FALSE, barcodes may be detected as blocks of some other type (e.g. pictures).
DetectInvertedImage	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine to detect whether the image is inverted (white text against black background). The text color is detected during layout analysis, and if it differs from normal, ABBYY FineReader Engine automatically inverts the image.

DetectOrientation	TRUE, **FALSE**	If this key is TRUE, the page orientation is detected during layout analysis, and if it differs from normal, ABBYY FineReader Engine automatically rotates the image.
DetectPictures	**TRUE**, FALSE	If this key is TRUE, the pictures are detected during layout analysis.
DetectTables	**TRUE**, FALSE	If this key is TRUE, the tables are detected during layout analysis.
FastObjectsExtraction	TRUE, **FALSE**	If this key is TRUE, layout analysis will speed up, but its quality may deteriorate. The value of this property is only relevant if ProhibitModelAnalysis is TRUE, otherwise it is ignored.
FlexiFormsDA	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine to locate all text on the page, including small text areas of low quality and text in diagrams and pictures. Tables are recognized as plain text. Note: If set to TRUE, FullTextIndexDA is automatically set to FALSE.
FullTextIndexDA	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine to detect all text on an image, including text embedded into the image. Reading order is not changed to provide ability for further full text search. Note: If this property is set to TRUE, FlexiFormsDA is automatically set to FALSE.
ProhibitClockwiseRotatio	TRUE, **FALSE**	Disables clockwise page rotation when selecting the page orientation. Note: This key must not have the TRUE value if ProhibitCounterclockwiseRotation and ProhibitUpsidedownRotation are set to TRUE.
ProhibitColorImage	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine to use only black and white plane during page layout analysis. In this case detection quality of colored tables and pictures can get worse.
ProhibitCounterclockwiseRotation	TRUE, **FALSE**	Disables counterclockwise page rotation when selecting the page orientation. Note: This key must not have the TRUE value if ProhibitClockwiseRotation and ProhibitUpsidedownRotation are set to TRUE.
ProhibitModelAnalysis	**TRUE**, FALSE	If this key is FALSE, typical variants of page layout will be gone through during page analysis and the best variant will be selected, which can improve recognition quality. If the best variant of page layout cannot be selected, standard page layout analysis will be performed.
ProhibitUpsidedownRotation	TRUE, **FALSE**	Disables upside down page rotation when selecting the page orientation. Note: This key must not have the TRUE value if ProhibitClockwiseRotation and ProhibitCounterclockwiseRotation are set to TRUE.
RemoveTexture	**TRUE**, FALSE	This key set to TRUE tells ABBYY FineReader Engine to remove the background noise from a temporary image used for recognition. The source image remains unaffected.
SingleColumnMode	TRUE, **FALSE**	If this key is set to TRUE, the analysis procedure presumes that there is only one column of text on a page.

Page Synthesis Parameters

Table: Page Synthesis Parameters		
Key	**Value**	**Description**
BalancedMode	TRUE, **FALSE**	If this key is TRUE, the recognition will run in balanced mode. The balanced mode is an intermediate mode between full and fast modes. The fast mode can be activated with the help of the FastMode key. This key is available for machine-printed texts only, for hand-printed texts the recognition will be run in full mode.
CorrectDynamicRange	TRUE, **FALSE**	If this key is TRUE, image colors will be corrected so that the background is white and the text is black, or vice versa, which improves image quality. Recognition, however, will slow down. We recommend using this key only if DetectBackgroundColor and DetectTextColor are TRUE.
DetectBackgroundColor	TRUE, **FALSE**	If this key is set to TRUE, the background color is detected during recognition.
DetectBold	**TRUE**, FALSE	If this key is set to TRUE, the bold-face type is detected during recognition.
DetectDropCaps	**TRUE**, FALSE	If this key is set to TRUE, the drop caps is detected during recognition.
DetectFontSize	**TRUE**, FALSE	If this key is set to TRUE, the font size is detected during recognition.
DetectItalic	**TRUE**, FALSE	If this key is set to TRUE, the italic-face type is detected during recognition.
DetectSerifs	**TRUE**, FALSE	If this key is set to TRUE, the serif typeface is detected during recognition.
DetectSmallCaps	**TRUE**, FALSE	If this key is set to TRUE, the small capital letters are detected during recognition.
DetectSubscripts-Superscripts	**TRUE**, FALSE	If this key is set to TRUE, the subscripts and superscripts are detected during recognition.
DetectTextColor	TRUE, **FALSE**	If this key is set to TRUE, the text color is detected during recognition.
DetectUnderline-Strikeout	**TRUE**, FALSE	If this key is set to TRUE, the underline and strikeout are detected during recognition.
InsertEmptyParagraphs-ForBigInterlines	TRUE,**FALSE**	If this key is set to TRUE, empty paragraphs are inserted to reproduce big line spacing of the original text. This key is ignored if ParagraphExtractionMode is NormalExtraction.
ExtractBlackSeparators	TRUE, **FALSE**	Specifies whether black separators should be searched during recognition. If this key is set to FALSE, the program will not look for black separators, which speeds up the entire recognition process.

FormatWithSpaces	TRUE, **FALSE**	Specifies whether space formatting should be performed instead of rich formatting (indents, tabs etc.). If this key is set to TRUE the ParagraphExtractionMode key is set to SingleLineParagraphsWithSpaceFormatting, the MonospaceDetectionMode key is set to Monospace, InsertEmptyParagraphsForBigInterlines is set to TRUE, and the DetectSerifs, DetectBold, DetectItalic, DetectUnderlineStrikeout, DetectSubscriptsSuperscripts, DetectSmallCaps, DetectDropcaps, DetectFontSize, DetectTextColor, DetectBackgroundColor keys are set to FALSE.
HighlightHyperlinks	TRUE, **FALSE**	If this key is set to TRUE, hyperlinks are identified by underlining and the color specified in the HyperlinksColor key.
HyperlinksColor	**0**	Specifies the hyperlinks color in decimal RGB format. The default is black (0).
KeepBullets	TRUE, **FALSE**	If this key is FALSE and the required bullet symbol is not found in the font, the program will make a search and substitute the symbol with another similar-looking symbol. If the key is TRUE, the symbol will not be substituted.
MonospaceDetectionMode	**Auto**, NotMonospace, Monospace	Specifies the mode of monospaced font detection.
ParagraphExtractionMode	**NormalExtraction**, RoughExtraction, SingleLineParagraphsWithSpaceFormatting, SingleLineParagraphsWithWordSeparationOnly	Specifies the mode of paragraph extraction.
RecognizedTextFontName	Serif: <FontName>, SansSerif: <FontName>, Monospace: <FontName>	Specifies the font names used in recognized text for different font types. The defaults are: Serif:TimeNewRoman, SansSerif:Arial, Monospace:CourierNew.
TextExtractionMode	**AutoDetect**, Both, RecognitionOnly, PdfInfoOnly	This key specifies the mode of PDF files recognition. This key is only relevant if the input file is in PDF format.

Recognizer Parameters

Table: Recognizer Parameters

Key	Value	Description
BalancedMode	TRUE, **FALSE**	If this key is TRUE, the recognition will run in balanced mode. The balanced mode is an intermediate mode between full and fast modes. The fast mode can be activated with the help of the FastMode key. This key is available for machine-printed texts only, for hand-printed texts the recognition will be run in full mode.

CaseRecognitionMode	**AutoCase**, SmallCase, CapitalCase	This key specifies the mode of letter case recognition. By default the value of this key is AutoCase, which corresponds to automatic case recognition.
CellsCount	**1**	Specifies the number of character cells for a recognized block. This key is valid only for the handprint recognition. It has a sense only for the field marking types that imply splitting the text in cells. Default value for this property is 1, but you should set the appropriate value to recognize the text correctly.
CJKTextDirection	**Autodetect**, Horizontal, Vertical	Sets the direction of the text to be recognized. This key is valid only for the hieroglyphic languages.
DisableSecondStageRecogn	TRUE, **FALSE**	If this key is set to TRUE, no second-stage recognition will be performed. The second stage of recognition is optional during recognition of small areas on the image. This function is needed to speed up recognition process. If this stage is skipped during recognition of full-page images, however, recognition quality may get worse.
ExactConfidenceCalculation	TRUE, **FALSE**	If this key is TRUE, character confidence will be defined more accurately, but recognition speed may get slower. This property is automatically set to TRUE if the SaveCharacterRecognitionVariants or SaveWordRecognitionVariants key is TRUE.
FastMode	TRUE, **FALSE**	This key set to TRUE provides 2-2.5 times faster recognition speed at the cost of a moderately increased error rate (1.5-2 times more errors). This property is available both for machine- and hand-printed texts. In the case of a hand-printed text, a special recognition mode is used. On good print quality texts, ABBYY FineReader Engine makes an average of 1-2 errors per page, and such moderate increase in error rate can be easily tolerated in many cases, such as full text indexing with "fuzzy" searches, preliminary recognition, etc. Note: We do not recommend using this mode to recognize small image fragments (for example, fragments which consist of only one line or word) because the time advantage will be insignificant.
FieldMarkingType	**SimpleText**, UnderlinedText, TextInFrame, GreyBoxes, CharBoxSeries, SimpleComb, CombInFrame, PartitionedFrame	This key specifies the type of marking around letters (for example, underline, frame, box and etc.). This key is valid only for the handprint recognition. By default the value of this property is SimpleText, which means the plain text. Note: For correct handprint recognition use CellsCount key that allows you to set the number of character cells for a recognized block.
OneLinePerBlock	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine to presume that the text in block to which the current RecognizerParams object belongs contains no more than one string.

OneWordPerLine	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine to presume that no text line may contain more than one word, so the lines of text will be recognized as a single word.
PossibleTextTypes	**Normal**, Typewriter, Matrix, Index, Handprinted, OCR_A, OCR_B, MICR_E13B, MICR_CMC7, Gothic,	This key contains a collection of Text-Type values. The key tells ABBYY FineReader Engine to presume that the text to recognize is of one of the types the collection contains. If the value of the TextType key is not ToBeDetected, the value of this property will be ignored. The collection should contain at least one element. When this key is changed, the TextType key is automatically set to ToBeDetected.
ProhibitHyphenation	TRUE, **FALSE**	This key set to TRUE prohibits recognition of hyphenation from line to line. It is useful when a text with presumably no hyphenations is recognized, in which case it may speed up the recognition. If there exist any hyphenations in the recognized block, and this key is TRUE, the hyphenated words will be recognized incorrectly.
ProhibitInterblock-Hyphenation	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine to presume that text from one block cannot be carried over to the next block.
ProhibitItalic	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine not to recognize letters printed with italic -styled font. It is useful when a text with presumably no italic letters is recognized, in which case it may speed up the recognition. If there exist any italic letters on the image, and this key is TRUE, these letters will be recognized incorrectly.
ProhibitSubscript	TRUE, **FALSE**	This key set to TRUE tells ABBYY FineReader Engine not to recognize superscript letters. It is useful when a text with presumably no superscripts is recognized, in which case it may speed up the recognition. If there exist any superscript letters on the image, and this property is TRUE, these letters will be recognized incorrectly.
ProhibitSuperscript	TRUE, **FALSE**	Specifies whether the variants of a character's recognition are saved.
SaveCharacter-RecognitionVariants	TRUE, **FALSE**	Specifies whether the variants of a character's recognition are saved.
SaveWordRecognition-Variants	TRUE, **FALSE**	Specifies whether the variants of recognition of a word are saved.
TextType	**Normal**, Typewriter, Matrix, Index, Handprinted, OCR_A, OCR_B, MICR_E13B, MICR_CMC7, Gothic, ToBeDetected	This key tells ABBYY FineReader Engine to presume that the text to recognize is of that type. By default the value of this property is Normal, which corresponds to a common typographic text.

`WritingStyle`	**Default**, American, German, Russian, Polish, Thai, Japanese, Arabic, Baltic, British, Bulgarian, Canadian, Czech, Croatian, French, Greek, Hungarian, Italian, Romanian, Slovak, Spanish, Turkish, Ukrainian, Common	Provides additional information about handprinted letters writing style.

B ABBYY FineReader 10

The main difference between ABBYY FR 8 and 10 is this: The ABBYY FineReader 8 works with parameters. Even though some parameters, i.e. those which did not alter between versions 8 and 10 of the FineReader Engine, still work with the FineReader 10, the FineReader Engine 10 is controlled using profiles. Languages for ABBYY 10 are the same as for ABBYY 8, see chapter Languages.

B.1 Profiles for ABBYY 10

ABBYY 10 supports two types of profiles. A profile represents a configuration setting. There is a set of predefined profiles available. Alternatively custom profiles can be created.

Predefined Profiles

Supported are the following pre-defined profiles:

- *DocumentConversion_Accuracy* – for converting documents into editable formats, optimized for accuracy
- *DocumentConversion_Speed* – for converting documents into editable formats, optimized for speed
- *DocumentArchiving_Accuracy* –for creating an electronic archive, optimized for accuracy
- *DocumentArchiving_Speed* – for creating an electronic archive, optimized for speed
- *BookArchiving_Accuracy* –for creating an electronic library, optimized for accuracy
- *BookArchiving_Speed* –for creating an electronic library, optimized for speed
- *TextExtraction_Accuracy* –for extracting text from documents, optimized for accuracy
- *TextExtraction_Speed* –for extracting text from documents, optimized for speed
- *FieldLevelRecognition* –for recognizing short text fragments
- *BarcodeRecognition* –for extracting barcodes
- *Version9Compatibility* –provided for compatibility, sets the processing parameters to the default values of ABBYY FineReader Engine 9.0.

Example: (Shell) Set the predefined profile DocumentArchiving_Accuracy.

```
-ocr "abbyy10" -ocp "PredefinedProfile = DocumentArchiving_Accuracy"
```

Custom Profiles

A custom profile is defined by a configuration file that is set up like an ini file. It consists of sections and entries in each section. The following sections are supported:

- PrepareImageMode
- PageProcessingParams
- PageAnalysisParams
- BarcodeParams

- RecognizerParams
- ObjectsExtractionParams
- OrientationDetectionParams
- SynthesisParamsForDocument

They are described in the chapter ABBYY 10 Parameters. Here is an example of a custom profile file:

Example:

```
[PrepareImageMode]
DiscardColorImage = false
[RecognizerParams]
BalancedMode = false
TextLanguage = English,German
```

The custom profile is provided using the parameter “Profile”. Ensure the profile file resides at a location that can be accessed by the OCR engine.

Example: (Shell) Set the custom profile Profile1.ini.

```
-ocr "abbyy10" -ocp "Profile=C:\path\Profile1.ini"
```

B.2 ABBYY 10 Parameters

Prepare Image Mode

`[PrepareImageMode]`

Table: Prepare Image Mode		
`AutoOverwrite-Resolution`	Boolean	Specifies whether resolution of the prepared image should be automatically overwritten. The property is only available, if the value of the OverwriteResolution property is FALSE. If the value of the AutoOverwriteResolution property is TRUE, ABBYY FineReader Engine will automatically detect and overwrite image resolution. By default, the value of the property is TRUE.

`CorrectSkewBy-BlackSquaresHorizontally`
`CorrectSkewBy-BlackSquaresVertically`
`CorrectSkewBy-HorizontalLines`
`CorrectSkewBy-HorizontalText`
`CorrectSkewBy-VerticalLines`
`CorrectSkewBy-VerticalText`
	Boolean	These properties are obsolete. Use the CorrectSkewMode property instead.

`CorrectSkewMode`	Long	Specifies the mode of skew correction. The value of this property is an OR superposition of the CorrectSkewModeEnum enumeration constants which denote the types of skew correction. 0 means do not correct skew. By default, this property is set to CSM_CorrectSkewByHorizontalText | CSM_CorrectSkewByVerticalText.

CreatePreview	Boolean	This property set to TRUE tells ABBYY FineReader Engine to create preview page for the prepared image. By default, this property is set to FALSE.
DiscardColorImage	Boolean	This property set to TRUE tells ABBYY FineReader Engine to leave only black-and-white planes in the prepared image. By default, this property is set to FALSE.
ImageCompression	ImageCompressionEnum	This property specifies how an image should be compressed during conversion to the internal format. By default, this property is set to IC_Auto.
InvertImage	Boolean	This property set to TRUE tells ABBYY FineReader Engine to invert colors of the prepared image. By default, this property is set to FALSE.
MirrorImage	Boolean	This property set to TRUE tells ABBYY FineReader Engine to mirror the prepared image around its vertical axis. By default, this property is set to FALSE.
OverwriteResolution	Boolean	Allows you to overwrite resolution of the prepared image. The resolution is overwritten depending on the values of the XResolutionToOverwrite and YResolutionToOverwrite properties. In this case the new resolution will be used for image preprocessing (i.e. for binarization, deskewing, etc.). Image resolution can be automatically overwritten (see the description of the AutoOverwriteResolution property). By default, this property is set to FALSE. See also IImageDocument::ChangeResolution.
PreviewHeight	Long	Specifies the height in pixels of the preview page. This property is valid only if the CreatePreview property is TRUE, otherwise it is ignored. By default, this property is set to 90.
PreviewWidth	Long	Specifies the width in pixels of the preview page. This property is valid only if the CreatePreview property is TRUE, otherwise it is ignored. By default, this property is set 64.
Rotation	RotationTypeEnum	This property specifies the rotation angle to apply to the image during preparation. It specifies no rotation by default. Support enums are: RT_UnknownRotation, RT_NoRotation, RT_Clockwise, RT_Counterclockwise, RT_Upsidedown

XResolutionToOverwrite	Long	Specifies the horizontal resolution of the original image in dpi. This value is used to overwrite resolution of the prepared image when resolution of the original image is not specified or incorrect and only if the OverwriteResolution property is TRUE. ABBYY FineReader Engine works with the prepared image which horizontal and vertical resolutions are equal, therefore the program stretches the image so that the horizontal and vertical resolutions of the prepared image are identical and equal to the maximum of ResolutionToOverwrite and YResolutionToOverwrite. By default, this property is set to 300.
YResolutionToOverwrite	Long	Specifies the vertical resolution of the original image in dpi. This value is used to overwrite resolution of the prepared image when resolution of the original image is not specified or incorrect and only if the OverwriteResolution property is TRUE. ABBYY FineReader Engine works with the prepared image which horizontal and vertical resolutions are equal, therefore the program stretches the image so that the horizontal and vertical resolutions of the prepared image are identical and equal to the maximum of XResolutionToOverwrite and YResolutionToOverwrite. By default, this property is set to 300.

Page Processing Parameters

[PageProcessingParams]

Table: Page Processing Parameters		
DetectBarcodes	Boolean	Specifies if barcodes are detected, and accordingly barcode blocks created, during page processing. If this property is FALSE, barcodes may be detected as blocks of some other type (e.g. pictures). This property is FALSE by default.
DetectOrientation	Boolean	If this property is TRUE, the page orientation is detected during page processing, and if it differs from normal, ABBYY FineReader Engine automatically rotates the image. This property is FALSE by default.
PerformPageAnalysis	Boolean	Specifies if page analysis is to be performed. If this property is FALSE, the PageAnalysisParams property is ignored. This property is TRUE by default.
RemoveGeometrical-Distortions	Boolean	Specifies if geometrical distortions (perspective on photos, curved lines from scanned books, etc.) should be removed during layout analysis. This property is FALSE by default.

Page Analysis

[PageAnalysisParams]

Table: Page Analysis Parameters		
DetectPictures	Boolean	If this property is TRUE, the pictures are detected during layout analysis. This property is TRUE by default.
DetectSeparators	Boolean	If this property is TRUE, the separators are detected during layout analysis. This property is TRUE by default.
DetectVectorGraphics	Boolean	If this property is TRUE, the vector pictures are detected during layout analysis. Vector picture blocks may appear in the layout only if this property has been set to TRUE during layout analysis. This property is TRUE by default.
DetectTables	Boolean	If this property is TRUE, the tables are detected during layout analysis. This property is TRUE by default.
NoShadowsMode	Boolean	This property set to TRUE tells ABBYY FineReader Engine to presume that an image has no shadows from scanning. This property is FALSE by default.
ProhibitDoublePageMode	Boolean	This property set to TRUE tells ABBYY FineReader Engine to presume that an image is not a book double page. This property is FALSE by default.
ProhibitModelAnalysis	Boolean	If this property is FALSE, typical variants of page layout will be gone through during page analysis and the best variant will be selected, which can improve recognition quality. If the best variant of page layout cannot be selected, standard page layout analysis will be performed. This property is FALSE by default.
SingleColumnMode	Boolean	If this property is set to TRUE, the analysis procedure presumes that there is only one column of text on a page. This property is FALSE by default. The value of this property is ignored, if the ProhibitModelAnalysis property is set to FALSE.

Barcode Parameters

[BarcodeParams]

Table: Barcode Parameters		
HasChecksum	Boolean	Specifies whether the barcode being recognized must be interpreted as the barcode of the same type but with a check sum. This property is only available for barcodes of types Code 39, Interleaved 2 of 5, Codabar, and Matrix 2 of 5. By default, this property is set to FALSE. Note: While Codabar has no check digit, ABBYY FineReader Engine uses an algorithm for computing check digits according to Modulo 16. The check digit is computed as follows. Each Codabar character has a value assigned to it. The sum of all character values is taken, including the Start and the Stop characters. The data character whose value, when added to this sum, equals a multiple of 16 is the check digit.
IsCode39WithoutAsterisk	Boolean	Specifies that the Code 39 barcode being recognized has no start and stop symbol, the asterisk "*". By default, this property is set to FALSE.
Orientation	Long	The value of this property is an OR superposition of the BarcodeOrientationEnum enumeration constants which denote the types of barcode orientation. For example, if it is set to BO_Left_To_Right | BO_Down_To_Top, ABBYY FineReader Engine will presume that barcode blocks may be oriented either from left to right or from down to top, ignoring all other variants. By default, this property is set to BO_Autodetect, i.e. ABBYY FineReader Engine will detect the barcode orientation automatically.
PDF417CodePage	CodePageEnum	This property is used to recognize barcodes which do not conform to the barcode specifications. Do not use this property for barcodes created in conformity with the barcode specifications. Some barcode printers use code pages other than US-MSDOS required by the specifications. In this case, use this property to specify the code page which was used by the barcode printer to create the barcode. In most cases this will be the code page of the operating system under which the barcode printer was running. By default, this property is set to CP_Null.
SupplementType	Long	The value of this property is an OR superposition of the BarcodeSupplementTypeEnum enumeration constants. This property is only available for barcodes of the EAN 8, 13, UPC-A, and UPC-E types. For example, if it is set to BS_Void | BS_2Digits, ABBYY FineReader Engine will try to recognize barcode blocks either without supplementary barcode or with 2-digit supplementary barcode. By default, this property is set to BS_Autodetect, i.e. ABBYY FineReader Engine will detect the supplementary barcode type automatically.

Type	Long	The value of this property is an OR superposition of the BarcodeTypeEnum enumeration constants which denote the types of barcodes. For example, if it is set to BT_EAN13 | BT_EAN8, ABBYY FineReader Engine will try to recognize barcode blocks in either EAN 13 or EAN 8 standard, ignoring all other variants. By default, this property is set to BT_Autodetect, i.e. ABBYY FineReader Engine will detect the barcode type automatically.

Recognizer Parameters

[RecognizerParams]

Table: Recognizer Parameters

BalancedMode	Boolean	If this property is TRUE, the recognition will run in balanced mode. The balanced mode is an intermediate mode between full and fast modes. The fast mode can be activated with the help of the FastMode property. This property is available for machine-printed texts only, for hand-printed texts the recognition will be run in full mode. By default, this property is FALSE.
CaseRecognitionMode	CaseRecognitionModeEnum	This property specifies the mode of letter case recognition. By default the value of this property is CRM_AutoCase, which corresponds to automatic case recognition.
CellsCount	Long	Specifies the number of character cells for a recognized block. This property is valid only for the handprint recognition. It has a sense only for the field marking types (the FieldMarkingType property) that imply splitting the text in cells. Default value for this property is 1, but you should set the appropriate value to recognize the text correctly.
CJKTextDirection	CJKTextDirectionEnum	Sets the direction of the text to be recognized. This property is valid only for the hieroglyphic languages. By default, this property is CJKTD_Autodetect.
ErrorHiliteLevel	ErrorHiliteLevelEnum	Specifies the level at which the CharParams::IsSuspicious property is set to TRUE for a recognized character. The name of the property reflects the fact that the uncertain characters are highlighted with color in ABBYY FineReader. By default the value of this property is EHL_Standard.
ExactConfidenceCalculation	Boolean	If this property is TRUE, character and word confidence will be defined more accurately, but recognition speed may get slower. The value of character confidence is stored in the CharConfidence property of the CharacterRecognitionVariant and PlainText objects. The value of word confidence is stored in the WordConfidence property of the WordRecognitionVariant object. This property is automatically set to TRUE if the SaveCharacterRecognitionVariants or SaveWordRecognitionVariants property is TRUE. By default, this property is FALSE.

FastMode	Boolean	This property set to TRUE provides 2-2.5 times faster recognition speed at the cost of a moderately increased error rate (1.5-2 times more errors). This property is available both for machine- and hand-printed texts. In the case of a hand-printed text (text type TT_Handprinted), a special recognition mode is used. On good print quality texts, ABBYY FineReader Engine makes an average of 1-2 errors per page, and such moderate increase in error rate can be easily tolerated in many cases, such as full text indexing with "fuzzy" searches, preliminary recognition, etc. By default, this property is FALSE. Note: We do not recommend using this mode to recognize small image fragments (for example, fragments which consist of only one line or word) because the time advantage will be insignificant.
FieldMarkingType	FieldMarking-TypeEnum	This property specifies the type of marking around letters (for example, underline, frame, box, etc.). This property is valid only for the handprint recognition. By default the value of this property is FMT_SimpleText, which means the plain text. Note: For correct handprint recognition use CellsCount property that allows you to set the number of character cells for a recognized block.
LowResolutionMode	Boolean	Specifies whether a text on an image with low resolution is recognized. By default, the value of this property is FALSE.
OneLinePerBlock	Boolean	This property set to TRUE tells ABBYY FineReader Engine to presume that the text in block to which the current RecognizerParams object belongs contains no more than one string. By default this property is FALSE.
OneWordPerLine	Boolean	This property set to TRUE tells ABBYY FineReader Engine to presume that no text line may contain more than one word, so the lines of text will be recognized as a single word. By default this property is FALSE. The property is obsolete. Use the TextTypes property instead.
PossibleTextTypes	LongsCollection	This property contains a collection of TextTypeEnum values. The property tells ABBYY FineReader Engine to presume that the text to recognize is of one of the types the collection contains. If the value of the TextType property is not TT_ToBeDetected, the value of this property will be ignored. The property returns a copy of the collection but not a reference to it. In order to modify the value of the property it is necessary to create a new collection, add required values to it, and then assign the collection to the property. The collection should contain at least one element and cannot contain TT_ToBeDetected. When this property is changed, the TextType property is automatically set to TT_ToBeDetected. By default it contains TT_Normal.

ProhibitHyphenation	Boolean	This property set to TRUE prohibits recognition of hyphenation from line to line. It is useful when a text with presumably no hyphenations is recognized, in which case it may speed up the recognition. If there exist any hyphenations in the recognized block, and this property is TRUE, the hyphenated words will be recognized incorrectly. By default this property is FALSE.
ProhibitInterblockHyphen	Boolean	This property set to TRUE tells ABBYY FineReader Engine to presume that text from one block cannot be carried over to the next block. By default this property is FALSE.
ProhibitItalic	Boolean	This property set to TRUE tells ABBYY FineReader Engine not to recognize letters printed with italic -styled font. It is useful when a text with presumably no italic letters is recognized, in which case it may speed up the recognition. If there exist any italic letters on the image, and this property is TRUE, these letters will be recognized incorrectly. By default this property is FALSE.
ProhibitSubscript	Boolean	This property set to TRUE tells ABBYY FineReader Engine not to recognize subscript letters. It is useful when a text with presumably no subscripts is recognized, in which case it may speed up the recognition. If there exist any subscript letters on the image, and this property is TRUE, these letters will be recognized incorrectly. By default this property is FALSE.
ProhibitSuperscript	Boolean	This property set to TRUE tells ABBYY FineReader Engine not to recognize superscript letters. It is useful when a text with presumably no superscripts is recognized, in which case it may speed up the recognition. If there exist any superscript letters on the image, and this property is TRUE, these letters will be recognized incorrectly. By default this property is FALSE.
SaveCharacter-RecognitionVariants	Boolean	Specifies whether the variants of characters recognition are saved. The ICharParams::CharacterRecognitionVariants property returns a collection of recognition variants for a character. The default value is FALSE. See also Using Voting API.
SaveCharacterRegions	Boolean	Specifies whether the exact characters regions (ICharParams::CharacterRegion) are saved. The default value is FALSE.
SaveWordRecognition-Variants	Boolean	Specifies whether the variants of recognition of a word are saved. The IParagraph::GetWordRecognitionVariants method and ICharParams::WordRecognitionVariants property return a collection of recognition variants for a word. The default value is FALSE. See also Using Voting API.
TextLanguage	TextLanguage	This property refers to the TextLanguage object used for image recognition. By default this parameter is initialized with English language. This property may be easily set via the SetPredefinedTextLanguage method.

TextTypes	Long	This property tells ABBYY FineReader Engine to presume that the text to recognize is of that type. By default the value of this property is TT_Normal. Note: If this property is set to TT_ToBeDetected, TT_Handprinted, or TT_Index, the TrainUserPatterns property cannot be set to TRUE. The value of this property is an OR superposition of the TextTypeEnum enumeration constants which denote possible text types used for recognition. For example, if it is set to TT_Normal | TT_Index, ABBYY FineReader Engine will presume that the text contains only common typographic text and digits written in ZIP-code style, ignoring all other variants. By default, this property is set to TT_Normal. The property cannot be set to TT_ToBeDetected. See also Using Text Type Autodetection. Notes: If this property is set to TT_Handprinted, or TT_Index, the TrainUserPatterns property cannot be set to TRUE. If this property is equal to any combination of TT_Matrix, TT_Typewriter, TT_OCR_A, and TT_OCR_B, italic fonts and superscript/subscript will not be recognized, regardless of the values of the ProhibitItalic, ProhibitSubscript and ProhibitSuperscript properties.
TrainUserPatterns	Boolean	This property specifies whether user patterns should be trained during the recognition. If this property is TRUE, some user pattern file should be specified in the UserPatternsFile property. The Pattern Training dialog box will display during recognition. For correct operation of pattern training process it is necessary to set the value of the parent window HWND handle (IEngine::ParentWindow property). See also Recognizing with Training. By default this property is FALSE. If this property is set to TRUE, the TextType and TextTypes properties cannot be set to TT_ToBeDetected, TT_Handprinted, or TT_Index. Notes: Pattern training is not supported for hieroglyphic languages.
UseBuiltInPatterns	Boolean	This property set to TRUE means that ABBYY FineReader Engine will use its own built-in patterns for recognition. Patterns are files establishing relationship between character image and character itself. By default this property is TRUE. You may want to set this property to FALSE when you do not want to use standard ABBYY FineReader Engine patterns for character recognition, but user patterns only. This may be useful for recognition of text typed with decorative or non- standard fonts. In this case it is better not to use ABBYY FineReader Engine built-in patterns, but use your own user-defined patterns trained for these fonts.

UserPatternsFile	String	UserPatternsFile property. If the UserPatternsFile property is empty the UseBuiltInPatterns property is ignored. See also Recognizing with Training. Contains the full path to a file of the user pattern used for recognition. By default this property stores an empty string. If the value of this property is not empty, information from the user pattern file will be used during recognition. If the UseBuiltInPatterns property is FALSE, which means that standard ABBYY FineReader Engine patterns are not used during recognition, this property should contain a path to user-defined pattern file, as only information stored in it will be used. See also Recognizing with Training.
WritingStyle	WritingStyleEnum	Provides additional information about handprinted letters writing style. By default the value of this property is WS_Default, which means that the writing style is selected depending on the current language of the operating system.

Objects Extraction Parameters

[ObjectsExtractionParams]

Table: Objects Extraction Parameters		
DetectMatrixPrinter	Boolean	If this property is TRUE, the text printed on matrix printer is detected during objects extraction. This property is TRUE by default.
DetectPorousText	Boolean	If this property is TRUE, the regions with porous text are detected during objects extraction. This property is TRUE by default.
FastObjectsExtraction	Boolean	If this property is TRUE, objects extraction will speed up, but its quality may deteriorate. This property is FALSE by default.
FlexiFormsDA	Boolean	This property set to TRUE tells ABBYY FineReader Engine to locate all text on the page, including small text areas of low quality and text in diagrams and pictures. Tables are recognized as plain text. This property is FALSE by default.
FullTextIndexDA	Boolean	This property set to TRUE tells ABBYY FineReader Engine to detect all text on an image, including text embedded into the image. Reading order is not changed to provide ability for further full-text search. This property is FALSE by default.
ProhibitColorImage	Boolean	This property set to TRUE tells ABBYY FineReader Engine to use only black-and-white plane during objects extraction. In this case detection quality of colored tables and pictures can get worse. This property is FALSE by default.
RemoveGarbage	Boolean	Specifies if garbage (excess dots that are smaller than a certain size) is to be removed from the image during objects extraction. This property is FALSE by default.

RemoveTexture	Boolean	This property set to TRUE tells ABBYY FineReader Engine to remove the background noise from a temporary image used for recognition. The source image remains unaffected. This property is TRUE by default.

Orientation Detection Parameters

[OrientationDetectionParams]

Table: Orientation Detection Parameters

OrientationDetectionMode	OrientationDetectionModeEnum	Specifies the mode of page orientation detection. Supported enums are: ODM_Fast, ODM_Normal, ODM_Thorough. This property is ODM_Normal by default.
ProhibitClockwiseRotation	Boolean	Disables clockwise page rotation when selecting the page orientation. This property is FALSE by default. Note: This property must not have the TRUE value if the ProhibitCounterclockwiseRotation and ProhibitUpsidedownRotation properties are set to TRUE.
ProhibitCounterclockwiseRotation	Boolean	Disables counterclockwise page rotation when selecting the page orientation. This property is FALSE by default. Note: This property must not have the TRUE value if the ProhibitClockwiseRotation and ProhibitUpsidedownRotation properties are set to TRUE.
ProhibitUpsidedownRotati	Boolean	Disables upside-down page rotation when selecting the page orientation. This property is FALSE by default. Note: This property must not have the TRUE value if the ProhibitClockwiseRotation and ProhibitCounterclockwiseRotation properties are set to TRUE.

Synthesis Parameters

[SynthesisParamsForDocument]

Table: Synthesis Parameters		
DetectDocumentStructure	Boolean	Specifies whether document structure detection should be performed while document synthesis. This property is set to TRUE by default.
DetectFontFormatting	Boolean	Specifies whether font formatting detection should be performed while document synthesis. This property is set to TRUE by default.
InsertEmptyParagraphs-ForBigInterlines	Boolean	If this property is set to TRUE, empty paragraphs are inserted to reproduce big line spacing of the original text. This property is set to FALSE by default.
PagePoolSize	Long	Specifies how many pages may be loaded by document synthesis simultaneously. This property allows you to decrease memory usage. We recommend to use the value in range from 32 to 64. The more the value, the more speed of processing. However, for processing big documents it is not recommended to use the highest values of this property, as this may lead to an out of memory error. The value less than 5 is ignored. By default the value of this property is 64.
SaveRecognitionInfo	Boolean	SaveCharacterRecognitionVariants property or the SaveWordRecognitionVariants property of the RecognizerParams object is set to TRUE, the value of this property is ignored. This property is set to TRUE by default.

C ABBYY FineReader 11

Between version 10 and 11, there are no fundamental changes.

Note: Some properties from ABBYY 10 have been removed or renamed. Those are listed in the documentation of the respective sections.

C.1 Languages for ABBYY 11

The following languages are supported:

C.1.1 Natural languages

Abkhaz – Abkhaz
Adyghe – Adyghe
Afrikaans – Afrikaans
Agul – Agul
Albanian – Albanian
Altaic – Altaic
Arabic – Arabic (Saudi Arabia)
ArmenianEastern – Armenian (Eastern)
ArmenianGrabar – Armenian (Grabar)
ArmenianWestern – Armenian (Western)
Awar – Avar
Aymara – Aymara
AzeriCyrillic – Azerbaijani (Cyrillic)
AzeriLatin – Azerbaijani (Latin)
Bashkir – Bashkir
Basque – Basque
Belarusian – Belarussian
Bemba – Bemba
Blackfoot – Blackfoot
Breton – Breton

Bugotu – Bugotu
Bulgarian – Bulgarian
Buryat – Buryat
Catalan – Catalan
Chamorro – Chamorro
Chechen – Chechen
ChinesePRC – Chinese Simplified
ChineseTaiwan – Chinese Traditional
Chukcha – Chukcha
Chuvash – Chuvash
Corsican – Corsican
CrimeanTatar – Crimean Tatar
Croatian – Croatian
Crow – Crow
Czech – Czech
Danish – Danish
Dargwa – Dargwa
Dungan – Dungan
Dutch – Dutch (Netherlands)
DutchBelgian – Dutch (Belgium)
English – English
EskimoCyrillic – Eskimo (Cyrillic)
EskimoLatin – Eskimo (Latin)
Esperanto – Esperanto
Estonian – Estonian
Even – Even
Evenki – Evenki
Faeroese – Faeroese
Farsi – Farsi
Fijian – Fijian
Finnish – Finnish
French – French
Frisian – Frisian
Friulian – Friulian
GaelicScottish – Scottish Gaelic
Gagauz – Gagauz
Galician – Galician
Ganda – Ganda
German – German
GermanNewSpelling – German (new spelling)
GermanLuxembourg – German (Luxembourg)
Greek – Greek
Guarani – Guarani
Hani – Hani
Hausa – Hausa
Hawaiian – Hawaiian
Hebrew – Hebrew
Hungarian – Hungarian
Icelandic – Icelandic
Ido – Ido
Indonesian – Indonesian
Ingush – Ingush
Interlingua – Interlingua
Irish – Irish
Italian – Italian
Japanese – Japanese
Kabardian – Kabardian
Kalmyk – Kalmyk
KarachayBalkar – Karachay-Balkar
Karakalpak – Karakalpak
Kasub – Kasub
Kawa – Kawa
Kazakh – Kazakh
Khakas – Khakas
Khanty – Khanty
Kikuyu – Kikuyu
Kirgiz – Kirghiz
Kongo – Kongo
Korean – Korean
KoreanHangul – Korean (Hangul)
Koryak – Koryak
Kpelle – Kpelle
Kumyk – Kumyk
Kurdish – Kurdish
Lak – Lak
Lappish – Sami (Lappish)
Latin – Latin
Latvian – Latvian
LatvianGothic – Latvian language written in Gothic script
Lezgin – Lezgin
Lithuanian – Lithuanian
Luba – Luba
Macedonian – Macedonian
Malagasy – Malagasy
Malay – Malay
Malinke – Malinke
Maltese – Maltese
Mansi – Mansi
Maori – Maori
Mari – Mari
Maya – Maya
Miao – Miao
Minankabaw – Minangkabau
Mohawk – Mohawk
Mongol – Mongol
Mordvin – Mordvin
Nahuatl – Nahuatl
Nenets – Nenets
Nivkh – Nivkh
Nogay – Nogay
Norwegian – NorwegianNynorsk and Norwegian-Bokmal
NorwegianBokmal – Norwegian (Bokmal)
NorwegianNynorsk – Norwegian (Nynorsk)
Nyanja – Nyanja
Occidental – Occidental
Ojibway – Ojibway

OldEnglish – Old English
OldFrench – Old French
OldGerman – Old German
OldItalian – Old Italian
OldSlavonic – Old Slavonic
OldSpanish – Old Spanish
Ossetic – Ossetian
Papiamento – Papiamento
PidginEnglish – Tok Pisin
Polish – Polish
PortugueseBrazilian – Portuguese (Brazil)
PortugueseStandard – Portuguese (Portugal)
Provencal Provencal
Quechua – Quechua
RhaetoRomanic – Rhaeto-Romanic
Romanian – Romanian
RomanianMoldavia – Romanian (Moldavia)
Romany – Romany
Ruanda – Ruanda
Rundi – Rundi
RussianOldSpelling – Russian (old spelling)
Russian – Russian
RussianWithAccent – Russian (with accents marking stress position)
Samoan – Samoan
Selkup – Selkup
SerbianCyrillic – Serbian (Cyrillic)
SerbianLatin – Serbian (Latin)
Shona – Shona
Sioux – Sioux (Dakota)
Slovak – Slovak
Slovenian – Slovenian
Somali – Somali
Sorbian – Sorbian
Sotho – Sotho
Spanish – Spanish
Sunda – Sunda
Swahili – Swahili
Swazi – Swazi
Swedish – Swedish
Tabassaran – Tabassaran
Tagalog – Tagalog
Tahitian – Tahitian
Tajik – Tajik
Tatar – Tatar
Thai – Thai
Tinpo – Jingpo
Tongan – Tongan
Tswana – Tswana
Tun – Tun
Turkish – Turkish
Turkmen – Turkmen
TurkmenLatin – Turkmen (Latin)
Tuvin – Tuvan
Udmurt – Udmurt
UighurCyrillic – Uighur (Cyrillic)
UighurLatin – Uighur (Latin)
Ukrainian – Ukrainian
UzbekCyrillic – Uzbek (Cyrillic)
UzbekLatin – Uzbek (Latin)
Vietnamese – Vietnamese
Visayan – Cebuano
Welsh – Welsh
Wolof – Wolof
Xhosa – Xhosa
Yakut – Yakut
Yiddish – Yiddish
Zapotec – Zapotec
Zulu – Zulu

C.1.2 Technical

Basic – Basic programming language
C++ – C/C++ programming language
Chemistry – Simple chemical formulas
CMC7 – For MICR CMC-7 text type
Cobol – Cobol programming language
Digits – Numbers
E13B – For MICR (E-13B) text type
Fortran – Fortran programming language
Java – Java programming language
OcrA – For OCR-A text type
OcrB – For OCR-B text type
Pascal – Pascal programming language

C.2 Profiles for ABBYY 11

ABBYY 11 supports two types of profiles. A profile represents a configuration setting. There is a set of predefined profiles available. Alternatively custom profiles can be created.

C.2.1 Predefined Profiles

Supported are the following pre-defined profiles:

DocumentConversion_Accuracy – for converting documents into editable formats, optimized for accuracy
- Best quality. Enables font style detection.

DocumentConversion_Speed – for converting documents into editable formats, optimized for speed
- Like DocumentConversion_Accuracy, but document analysis and recognition are speeded up.
- *Important! This profile requires the Fast Mode module available in the license.*

DocumentArchiving_Accuracy – for creating an electronic archive, optimized for accuracy
- Enables detection of maximum text on an image, including text embedded into the image.
- Skew correction is not performed.
- Fonts and styles are not detected.

DocumentArchiving_Speed – for creating an electronic archive, optimized for speed
- Like DocumentArchiving_Accuracy, but document analysis and recognition are speeded up.
- *Important! This profile requires the Fast Mode module*

BookArchiving_Accuracy – for creating an electronic library, optimized for accuracy
- Best quality. Enables font style detection.

BookArchiving_Speed – for creating an electronic library, optimized for speed
- Like BookArchiving_Accuracy, but document analysis and recognition are speeded up.
- *Important! This profile requires the Fast Mode module available in the license.*

TextExtraction_Accuracy – for extracting text from documents, optimized for accuracy
- Enables detection of all text on an image, including small text areas of low quality (pictures and tables are not detected).
- Fonts and styles are not detected.

TextExtraction_Speed – for extracting text from documents, optimized for speed
- Like TextExtraction_Accuracy, but document analysis and recognition are speeded up.
- *Important! This profile requires the Fast Mode module available in the license.*

FieldLevelRecognition – for recognizing short text fragments.
- Currently this profile has default settings.

BarcodeRecognition_Accuracy – for barcode extraction, optimized for accuracy
- Extracts only barcodes (texts, pictures, or tables are not detected).
- *Important! This profile requires the Barcode Autolocation module available in the license.*

BarcodeRecognition_Speed – for barcode extraction, optimized for speed
- Like BarcodeRecognition_Accuracy, but document analysis and recognition are speeded up.
- *Important! This profile requires the Barcode Autolocation module available in the license.*

BusinessCardsProcessing – for recognizing business cards
- Detects only business cards.
- Enables detection of all text on an image, including small text areas of low quality (pictures and tables are not detected).
- Resolution correction is performed.

EngineeringDrawingsProcessing – for recognizing technical drawings
- It takes into account large size and complexity of engineering diagrams, as well as possibility of different text orientation within the image.
- Enables detection of all text on an image, including text blocks of vertical orientation.

Version9Compatibility – for compatibility
- sets the processing parameters to the default values of ABBYY FineReader Engine 9.0.

Default – for default values
- Sets all the processing parameters to the default values.

Example: (Shell) Set the predefined profile DocumentArchiving_Accuracy.

```
-ocr "abbyy11" -ocp "PredefinedProfile = DocumentArchiving_Accuracy"
```

C.2.2 Custom Profiles

A custom profile is defined by a configuration file that is set up like an ini file. It consists of sections and entries in each section. The following sections are supported:

- PrepareImageMode
- ImageProcessingParams
- DocumentProcessingParams
- PageProcessingParams
- PagePreprocessingParams
- PageAnalysisParams
- TableAnalysisParams
- BarcodeParams
- ObjectsExtractionParams
- OrientationDetectionParams
- RecognizerParams
- SynthesisParamsForPage
- SynthesisParamsForDocument
- FontFormattingDetectionParams

They are described in the chapter ABBYY 11 Parameters. Here is an example of a custom profile file:

Example:

```
[PrepareImageMode]
DiscardColorImage = false
[RecognizerParams]
BalancedMode = false
TextLanguage = English,German
```

The custom profile is provided using the parameter “Profile”. Ensure the profile file resides at a location that can be accessed by the OCR engine.

Example: (Shell) Set the custom profile Profile1.ini.

```
-ocr "abbyy10" -ocp "Profile=C:\path\Profile1.ini"
```

C.2.3 Profile Parameters

Note: Some properties from ABBYY 10 have been removed or renamed. Those are listed in the documentation of the respective sections.

Prepare Image Mode

```
[PrepareImageMode]
```

Main settings		
`Rotation`	RotationTypeEnum	This property specifies the rotation angle to apply to the image during preparation. Supported values are: RT_UnknownRotation, RT_NoRotation, RT_Clockwise, RT_Counterclockwise, RT_Upsidedown It specifies no rotation by default.

CorrectSkew	BOOL	This property set to TRUE tells ABBYY FineReader Engine to correct skew during image preparation. The type of skew correction is defined by the CorrectSkewMode property. If this property is set to FALSE, the value of the CorrectSkewMode property is ignored. By default, this property is set to TRUE.
CorrectSkewMode	int	Specifies the mode of skew correction. The value of this property is an OR superposition of the CorrectSkewModeEnum enumeration constants which denote the types of skew correction. 0 means do not correct skew. The value of this property is ignored, if the CorrectSkew property is set to FALSE. By default, this property is set to CSM_CorrectSkewByHorizontalText | CSM_CorrectSkewByVerticalText.
BackgroundFillingColor	int	Specifies the color used for filling the areas which are added to the image after skew correction. The default value of this property is -1, which means that the color is determined by ABBYY FineReader Engine automatically. *Note:* The int value is calculated from the RGB triplet using the formula: (red value) + (256 x green value) + (65536 x blue value), where red value is the first triplet component, green value is the second triplet component, blue value is the third triplet component. Hence the int value of the color white equals 16777215.
InvertImage	BOOL	This property set to TRUE tells ABBYY FineReader Engine to invert colors of the prepared image. By default, this property is set to FALSE.
MirrorImage	BOOL	This property set to TRUE tells ABBYY FineReader Engine to mirror the prepared image around its vertical axis. By default, this property is set to FALSE.
EnhanceLocalContrast	BOOL	Specifies whether the local contrast of the image should be increased. Such preprocessing may increase the quality of recognition. *Important!* Setting this property to TRUE is meaningful for color and gray images only. By default, this property is set to FALSE.
PhotoProcessingMode	PhotoProcessingModeEnum	Specifies whether the processing image should be treated as a photo. If a photo is processed, FineReader Engine uses special algorithms for such image on different stages of image processing. Supported values are: PPM_TreatAsPhoto, PPM_TreatAsNonPhoto, PPM_Auto By default, this property is PPM_Auto, which means that FineReader Engine should define automatically whether the image is a photo.

Resolution overwriting		
AutoOverwrite-Resolution	BOOL	Specifies whether resolution of the prepared image should be automatically overwritten. The property is only available, if the value of the OverwriteResolution property is FALSE. If the value of the AutoOverwriteResolution property is TRUE, ABBYY FineReader Engine will automatically detect and overwrite image resolution. *Note:* If you set this property to FALSE and resolution of the prepared image is too low (less than 50 dpi), too high (more than 3200 dpi), or undefined, you should set the correct values of the XResolutionToOverwrite and YResolutionToOverwrite properties, otherwise an error is returned. By default, the value of the property is TRUE.
OverwriteResolution	BOOL	Allows you to overwrite resolution of the prepared image. The resolution is overwritten depending on the values of the XResolutionToOverwrite and YResolutionToOverwrite properties. In this case the new resolution will be used for image preprocessing (i.e. for binarization, deskewing, etc.). Image resolution can be automatically overwritten (see the description of the AutoOverwriteResolution property). By default, this property is set to FALSE.
XResolutionToOverwrite	int	Specifies the horizontal resolution of the original image in dpi. This value is used to overwrite resolution of the prepared image when resolution of the original image is not specified or incorrect and only if the OverwriteResolution property is TRUE. ABBYY FineReader Engine works with the prepared image which horizontal and vertical resolutions are equal, therefore the program stretches the image so that the horizontal and vertical resolutions of the prepared image are identical and equal to the maximum of ResolutionToOverwrite and YResolutionToOverwrite. By default, this property is set to 300.
YResolutionToOverwrite	int	Specifies the vertical resolution of the original image in dpi. This value is used to overwrite resolution of the prepared image when resolution of the original image is not specified or incorrect and only if the OverwriteResolution property is TRUE. ABBYY FineReader Engine works with the prepared image which horizontal and vertical resolutions are equal, therefore the program stretches the image so that the horizontal and vertical resolutions of the prepared image are identical and equal to the maximum of XResolutionToOverwrite and YResolutionToOverwrite. By default, this property is set to 300.

Creating image planes		
DiscardColorImage	BOOL	This property set to TRUE tells ABBYY FineReader Engine to leave only black-and-white planes in the prepared image.In this case image binarization is performed during image preparation. By default, this property is set to FALSE.
UseFastBinarization	BOOL	If this property is TRUE, ABBYY FineReader Engine will use algorithms for fast image binarization. Binarization is performed either when the image is loaded (if the DiscardColorImage is TRUE during preparation), or later when a black-and-white image is necessary. This property set to TRUE will speed up binarization however its quality may deteriorate. By default, the value of this property is FALSE.
ImageCompression	ImageCompressionEnum	This property specifies how an image should be compressed during conversion to the internal format. Possible values are: IC_Auto, IC_Compress, IC_NoCompression By default, this property is set to IC_Auto.
CreatePreview	BOOL	This property set to TRUE tells ABBYY FineReader Engine to create preview page for the prepared image. By default, this property is set to FALSE.
PreviewHeight	int	Specifies the height in pixels of the preview page. This property is valid only if the CreatePreview property is TRUE, otherwise it is ignored. By default, this property is set to 90.
PreviewWidth	int	Specifies the width in pixels of the preview page. This property is valid only if the CreatePreview property is TRUE, otherwise it is ignored. By default, this property is set 64.

Image Processing Parameters

[ImageProcessingParams]

Image Processing Parameters		
InvertImage	BOOL	Specifies if the image colors in the block are inverted. If this property is TRUE, ABBYY FineReader Engine will invert the image of a block before recognition. This property is FALSE by default.
MirrorImage	BOOL	Specifies if the image in the block is mirrored around the vertical axis. If this property is TRUE, ABBYY FineReader Engine will mirror the image of a block before recognition. This property is FALSE by default.
RotationType	RotationTypeEnum	Specifies the orientation of a text in a block relative to the normal reading position. Supported values are: RT_NoRotation, RT_Clockwise, RT_Counterclockwise, RT_Upsidedown This property cannot be set to RT_UnknownRotation. This property is RT_NoRotation by default, which means that the orientation is normal.

Document Processing Parameters

[DocumentProcessingParams]

Document Processing Parameters		
PerformSynthesis	BOOL	Specifies if document synthesis is to be performed. If this property is FALSE, the SynthesisParamsForDocument property is ignored. This property is TRUE by default.

Page Processing Parameters

[PageProcessingParams]

Page Processing Parameters		
PerformPreprocessing	BOOL	Specifies if page preprocessing is to be performed. Page preprocessing is performed before page analysis and includes correction of orientation, inversion, geometrical distortions. If this property is FALSE, the PagePreprocessingParams property is ignored. This property is TRUE by default.
ProhibitColorObjects-AtProcessing	BOOL	Specifies if color objects must be filtered out on the image before layout analysis and recognition. If this property is set to FALSE, the ColorObjectsProhibitingParams property is ignored. The default value of this property is FALSE.
PerformAnalysis	BOOL	Specifies if page analysis is to be performed. If this property is FALSE, the PageAnalysisParams property is ignored. This property is TRUE by default. *Note:* If your license does not support the Analysis module, the default value of this property is FALSE.
PerformRecognition	BOOL	Specifies if recognition is to be performed. If this property is FALSE, the RecognizerParams property is ignored. This property is TRUE by default. *Note:* If your license does not support the Recognition module, the default value of this property is FALSE.

ATTENTION: Removed or renamed properties		
DetectInvertedImage DetectOrientation RemoveGeometricalDistort	(Re-)moved	Use the CorrectInvertedImage, CorrectOrientation, OrientationDetectionParams, GeometryCorrectionMode properties of the PagePreprocessingParams section. There is a new stage concerned with image preprocessing, and settings used at that stage were grouped into a separate object to make the structure of parameters objects clearer and more logical.
PerformPageAnalysis	Renamed	Use the PerformAnalysis property to enable layout analysis and PerformPreprocessing property to enable correction of page orientation, inversion, and geometrical distortions. The image preprocessing stage has been separated from page analysis and has its own set of parameters.
DetectBarcodes	(Re-)moved	Use the corresponding properties of the PageAnalysisParams section. Many processing parameters were grouped into sections to make the structure of parameter sections clearer and more logical.

Page Preprocessing Parameters

[PagePreprocessingParams]

Page Preprocessing Parameters		
CorrectInvertedImage	BOOL	This property set to TRUE tells ABBYY FineReader Engine to detect whether the image is inverted (white text against black background). The text color is detected during page prerocessing, and if it differs from normal, ABBYY FineReader Engine automatically inverts the image. This property is FALSE by default.
CorrectOrientation	BOOL	If this property is TRUE, the page orientation is detected during page prerocessing, and if it differs from normal, ABBYY FineReader Engine automatically rotates the image. *Note:* If this property is set to TRUE: ▪ the TextTypes property of the RecognizerParams object cannot be set to TT_Handprinted; ▪ the TrainUserPatterns property of the RecognizerParams object cannot be set to TRUE. This property is FALSE by default.
CorrectShadowsAndHighlights	ThreeStatePropertyValueEnum	If this property is set to TSPV_Yes, the image preprocessing will include correction of excessive shadows and highlighting to improve recognition quality. This property is designed for use with photographs only. Possible values are: TSPV_Auto, TSPV_No, TSPV_Yes The default value of this property is TSPV_Auto.
CorrectSkew	ThreeStatePropertyValueEnum	If this property is set to TSPV_Yes, image skew will be corrected during page preprocessing. The type of skew correction is defined by the CorrectSkewMode property. If this property is set to TSPV_No, the value of the CorrectSkewMode property is ignored. By default, this property is set to TSPV_Auto.
CorrectSkewMode	int	Specifies the mode of skew correction. The value of this property is an OR superposition of the CorrectSkewModeEnum enumeration constants which denote the types of skew correction. 0 means do not correct skew. The value of this property is ignored, if the CorrectSkew property is set to TSPV_No. Possible values are: CSM_CorrectSkewByBlackSquaresHorizontally, CSM_CorrectSkewByBlackSquaresVertically, CSM_CorrectSkewByHorizontalLines, CSM_CorrectSkewByHorizontalText, CSM_CorrectSkewByVerticalLines, CSM_CorrectSkewByVerticalText. By default, this property is set to CSM_CorrectSkewByHorizontalText | CSM_CorrectSkewByVerticalText.

GeometryCorrectionMode	Geometry-CorrectionMode-Enum	Specifies whether geometrical distortions (perspective on photos, curved lines from scanned books, etc.) should be removed during page preprocessing. Possible values are: GCM_Correct, GCM_DontCorrect, GCM_Auto. This property is GCM_Auto by default.
ResolutionCorrection-Mode	Resolution-CorrectionMode-Enum	Specifies whether resolution of the image should be corrected during page preprocessing. Possible values are: RCM_Correct, RCM_DontCorrect, RCM_Auto. This property is RCM_Auto by default, which means that if resolution of the image is suspicious, FineReader Engine should automatically detect image resolution and change it.

Page Analysis

[PageAnalysisParams]

Block detection settings		
DetectText	BOOL	If this property is TRUE, the text areas are detected during layout analysis. This property is TRUE by default.
EnableTextExtraction-Mode	BOOL	This property set to TRUE tells FineReader Engine that text blocks can be located anywhere on the page. Isolated text blocks are detected during layout analysis. Tables are not detected. Model analysis is not performed, as if ProhibitModelAnalysis property was set to TRUE. This property is FALSE by default.
DetectTables	BOOL	If this property is TRUE, the tables are detected during layout analysis. This property is TRUE by default.
AggressiveTable-Detection	BOOL	This property manages the table detection mode. If you set it to TRUE, FineReader Engine tries to find as many tables as possible on the page. This setting is recommended only for the documents which contain a lot of tables. This property is FALSE by default.
DetectBarcodes	BOOL	Specifies if barcodes are detected, and accordingly barcode blocks created, during layout analysis. If this property is FALSE, barcodes may be detected as blocks of some other type (e.g. pictures). This property is FALSE by default.
DetectSeparators	BOOL	If this property is TRUE, the separators are detected during layout analysis. This property is TRUE by default.
DetectPictures	BOOL	If this property is TRUE, the pictures are detected during layout analysis. This property is TRUE by default.

DetectVectorGraphics	BOOL	If this property is TRUE, the vector pictures are detected during layout analysis. Vector picture blocks may appear in the layout only if this property has been set to TRUE during layout analysis. This property is TRUE by default.

Additional settings		
DetectMultiple-BusinessCards	BOOL	Specifies whether a processing page can contain several business cards. By default the value of this property is FALSE.
NoShadowsMode	BOOL	This property set to TRUE tells ABBYY FineReader Engine to presume that an image has no shadows from scanning. This property is FALSE by default.
DetectVertical-EuropeanText	BOOL	This property set to TRUE makes FineReader Engine look for vertically oriented text. It applies to all languages other than CJK. For CJK languages, vertical text detection is managed by the ProhibitCJKColumns property. This property is FALSE by default.
ProhibitCJKColumns	BOOL	The text in CJK languages can be written vertically as well as horizontally. Setting this property to TRUE tells ABBYY FineReader Engine to ignore the possibility of vertical text and recognize the image with the supposition that all text is arranged horizontally. This property is valid only for CJK languages. This property is FALSE by default.
ProhibitDoublePageMode	BOOL	This property set to TRUE tells ABBYY FineReader Engine to presume that an image is not a book double page. This property is FALSE by default.
ProhibitModelAnalysis	BOOL	If this property is FALSE, typical variants of page layout will be gone through during page analysis and the best variant will be selected, which can improve recognition quality. If the best variant of page layout cannot be selected, standard page layout analysis will be performed. *Note:* If EnableTextExtractionMode property is set to TRUE, the value of this property is ignored and model analysis is not performed. This property is FALSE by default.

ATTENTION: Removed or renamed properties		
FlexiFormsDA	Behavior changed	This property is obsolete and will be removed in future versions. The same functionality is now provided by EnableAggressiveTextExtraction property and IPageAnalysisParams::EnableTextExtractionMode property. If you do not set this property specifically, its value will be ignored. If you do, both properties mentioned above become automatically set to the value you assigned to this property.

SingleColumnMode	Removed	This property is no longer supported in FineReader Engine recognition technologies.

Table Analysis Parameters

[TableAnalysisParams]

Table Analysis Parameters		
DetectCellsInversion	BOOL	If this property is TRUE, the cells inversion is detected during table block analysis. This property is TRUE by default.
DetectCellsOrientation	BOOL	If this property is TRUE, the cells orientation is detected during table block analysis. This property is TRUE by default.
SingleLinePerCell	BOOL	Set this property to TRUE if you only recognize tables with one line of text per each cell. The table layout will be analyzed more readily. This property is FALSE by default.
SplitOnlyBySeparators	BOOL	Set this property to TRUE if you only recognize tables with no hidden separators. The table layout will be analyzed more readily. This property is FALSE by default.

Barcode Parameters

[BarcodeParams]

Main settings		
Type	int	The value of this property is an OR superposition of the BarcodeTypeEnum enumeration constants which denote the types of barcodes. For example, if it is set to BT_EAN13 | BT_EAN8, ABBYY FineReader Engine will try to recognize barcode blocks in either EAN 13 or EAN 8 standard, ignoring all other variants. By default, this property is set to BT_Autodetect, i.e. ABBYY FineReader Engine will detect the barcode type automatically. *Note:* The default value allows detecting barcodes of all supported types. However, we recommend excluding IATA 2 of 5, Interleaved 2 of 5, and Matrix 2 of 5 barcode types unless you are certain that they will occur on your images. Barcodes of these types do not have a checksum and can therefore be mistakenly found on image areas that do not actually contain any barcodes.

Orientation	int	The value of this property is an OR superposition of the BarcodeOrientationEnum enumeration constants which denote the types of barcode orientation. For example, if it is set to BO_Left_To_Right | BO_Down_To_Top, ABBYY FineReader Engine will presume that barcode blocks may be oriented either from left to right or from down to top, ignoring all other variants. By default, this property is set to BO_Autodetect, i.e. ABBYY FineReader Engine will detect the barcode orientation automatically.
MinRatioToTextHeight	double	This property defines the minimal acceptable height of the barcode in relation to the average letters height. Use this setting if you need to allow detecting low barcodes. You can only assign a positive value to this property, or the default value -1. The upper limit to this property value is 2147483647 ($= 2^{31} - 1$). The default value of this property is -1, which means that ABBYY FineReader Engine will adjust this setting automatically.

Settings for certain barcode types		
CodePage	CodePageEnum	This property is used to recognize PDF417, Aztec, Data Matrix, QR Code, MaxiCode barcodes which do not conform to the barcode specifications. Do not use this property for barcodes created in conformity with the barcode specifications. Some barcode printers use code pages other than the one required by the specifications. In this case, use this property to specify the code page which was used by the barcode printer to create the barcode. In most cases this will be the code page of the operating system under which the barcode printer was running. *Note:* The following code pages required by the specifications: ▪ for PDF417 – DOS United States (437) (CP_US_MSDOS) ▪ for Aztec, DataMatrix, QR Code, MaxiCode – ISO Latin 1 (8859-1) (CP_Latin_ISO) The property is used to convert recognized data to a Unicode string. By default, this property is set to CP_Null, which means that the code page required by the specification should be used.
ContainsBinaryData	BOOL	This property makes sense only for PDF417, Aztec, Data Matrix, QR Code barcodes which encode some binary data. The property affects the way binary data is represented in the recognized text. If this property is set to TRUE, the binary data encoded in a barcode is saved as a sequence of hexadecimal values for corresponding bytes. If this property is FALSE, the binary data is translated to the Unicode string using the code page specified in the CodePage property. By default, the value of this property is FALSE.

EnableAdvancedExtraction	BOOL	This property is only valid for twodimensional barcodes. When you set this property to TRUE, the advanced mode of barcode extraction is turned on. More possibilities of barcodes location are considered; in particular, it enables FineReader Engine to find barcodes which are positioned at 45° to the page orientation. However, this setting slows down the processing. The default value of this property is FALSE.
HasChecksum	BOOL	Specifies whether the barcode being recognized must be interpreted as the barcode of the same type but with a check sum. This property is only available for barcodes of types Code 39, Interleaved 2 of 5, Codabar, and Matrix 2 of 5. *Note:* While Codabar has no check digit, ABBYY FineReader Engine uses an algorithm for computing check digits according to Modulo 16. The check digit is computed as follows. Each Codabar character has a value assigned to it. The sum of all character values is taken, including the Start and the Stop characters. The data character whose value, when added to this sum, equals a multiple of 16 is the check digit. By default, this property is set to FALSE.
IsCode39WithoutAsterisk	BOOL	Specifies that the Code 39 barcode being recognized has no start and stop symbol, the asterisk "*". By default, this property is set to FALSE.
SupplementType	int	The value of this property is an OR superposition of the BarcodeSupplementTypeEnum enumeration constants. This property is only available for barcodes of the EAN 8, 13, UPC-A, and UPC-E types. For example, if it is set to BS_Void | BS_2Digits, ABBYY FineReader Engine will try to recognize barcode blocks either without supplementary barcode or with 2-digit supplementary barcode. Possible values are: BS_Unknown, BS_Void, BS_2Digits, BS_5Digits, BS_Autodetect. By default, this property is set to BS_Autodetect, i.e. ABBYY FineReader Engine will detect the supplementary barcode type automatically.

ATTENTION: Removed or renamed properties		
PDF417CodePage	Renamed	The new name of the property is CodePage. This property can be used for barcodes of several types besides PDF417.

Objects Extraction Parameters

[ObjectsExtractionParams]

Common settings		
FastObjectsExtraction	BOOL	If this property is TRUE, objects extraction will speed up, but its quality may deteriorate. This property is FALSE by default.
ProhibitColorImage	BOOL	This property set to TRUE tells ABBYY FineReader Engine to use only black-and-white plane during objects extraction. In this case detection quality of colored tables and pictures can get worse. This property is FALSE by default.

Objects removing		
RemoveGarbage	BOOL	Specifies if garbage (excess dots that are smaller than a certain size) is to be removed from the image during objects extraction. This property is FALSE by default.
RemoveTexture	BOOL	This property set to TRUE tells ABBYY FineReader Engine to remove the background noise from a temporary image used for recognition. The source image remains unaffected. This property is TRUE by default.

Objects detection		
DetectMatrixPrinter	BOOL	If this property is TRUE, the text printed on matrix printer is detected during objects extraction. This property is TRUE by default.
DetectPorousText	BOOL	If this property is TRUE, the regions with porous text are detected during objects extraction. This property is TRUE by default.
DetectTextOnPictures	BOOL	This property set to TRUE tells ABBYY FineReader Engine to detect all text on an image, including text embedded into images. Reading order is not changed to provide ability for full-text search later. This property is FALSE by default.
EnableAggressiveText-Extraction	BOOL	If this property is TRUE, FineReader Engine will try to extract as much text on the image as possible. We recommend using this mode when the image contains some low-quality text. This processing mode may lead to mistaken interpretation of pictures as text or vertical rearranging of the horizontal text. This property is FALSE by default.
ProhibitDotted-Separators	BOOL	If this property is set to TRUE, ABBYY FineReader Engine presumes that the document does not contain dotted separators. This property may be useful, if you are sure that the document does not contain dotted separators, or if some data is incorrectly recognized as a dotted separator. This is a temporary property, it may be removed in a future version. This property is FALSE by default.

ATTENTION: Removed or renamed properties		
FlexiFormsDA	Behavior changed	This property is obsolete and will be removed in future versions. The same functionality is now provided by EnableAggressiveTextExtraction property and IPageAnalysisParams::EnableTextExtractionMode property. If you do not set this property specifically, its value will be ignored. If you do, both properties mentioned above become automatically set to the value you assigned to this property.
FullTextIndexDA	Behavior changed	This property is obsolete and will be removed in future versions. The same functionality is now provided by the DetectTextOnPictures property.
SourceContentReuseMode	Default value changed	The default value of this property is now CRM_Auto. This setting uses both the text and the image layer of the PDF file and ensures the highest quality of objects extraction.

Orientation Detection Parameters

[OrientationDetectionParams]

Orientation Detection Parameters		
OrientationDetectionMode	OrientationDetectionModeEnum	Specifies the mode of page orientation detection. Possible values are: ODM_Fast, ODM_Normal, ODM_Thorough. This property is ODM_Normal by default.
ProhibitClockwiseRotation	BOOL	Disables clockwise page rotation when selecting the page orientation. *Note:* This property must not have the TRUE value if the ProhibitCounterclockwiseRotation and ProhibitUpsidedownRotation properties are set to TRUE. This property is FALSE by default.
ProhibitCounterclockwiseRotation	BOOL	Disables counterclockwise page rotation when selecting the page orientation. *Note:* This property must not have the TRUE value if the ProhibitClockwiseRotation and ProhibitUpsidedownRotation properties are set to TRUE. This property is FALSE by default.
ProhibitUpsidedownRotation	BOOL	Disables upside-down page rotation when selecting the page orientation. *Note:* This property must not have the TRUE value if the ProhibitClockwiseRotation and ProhibitCounterclockwiseRotation properties are set to TRUE. This property is FALSE by default.

Recognizer Parameters

[RecognizerParams]

Main settings		
TextLanguage	TextLanguage	This property refers to the TextLanguage object used for image recognition. By default this parameter is initialized with English language.
LanguageDetectionMode	ThreeState-PropertyValue-Enum	Manages automatic language detection. When language autodetection is on, the recognition language is detected for each word in the text. It is selected from the list of languages specified in the TextLanguage property. Autodetection is intended to be used during recognition of documents the language of which is not known to you. If you know for certain that all the languages you specified are present in the document, autodetection is useless. Turn it off by setting this property to TSPV_No. Possible values are: TSPV_No, TSPV_Yes, TSPV_Auto. By default this property value is TSPV_Auto.
TextTypes	int	The value of this property is an OR superposition of the TextTypeEnum enumeration constants which denote possible text types used for recognition. For example, if it is set to TT_Normal | TT_Index, ABBYY FineReader Engine will presume that the text contains only common typographic text and digits written in ZIP-code style, ignoring all other variants. See also Using Text Type Autodetection. *Notes:* ▪ If this property is equal to any combination of TT_Matrix, TT_Typewriter, TT_OCR_A, and TT_OCR_B, italic fonts and superscript/subscript will not be recognized, regardless of the values of the ProhibitItalic, ProhibitSubscript and ProhibitSuperscript properties. ▪ If this property is set to TT_Handprinted, or TT_Index, the TrainUserPatterns property cannot be set to TRUE. ▪ If this property is TT_Handprinted, the CorrectOrientation property of the PagePreprocessingParams section cannot be set to TRUE. ▪ If the TrainUserPatterns property is set to TRUE, the TextTypes property must contain only one text type. By default, this property is set to TT_Normal.

Recognition speed		
BalancedMode	BOOL	If this property is TRUE, the recognition will run in balanced mode. The balanced mode is an intermediate mode between full and fast modes. The fast mode can be activated with the help of the FastMode property. This property is available for machine-printed texts only, for hand-printed texts the recognition will be run in full mode. By default, this property is FALSE.
FastMode	BOOL	This property set to TRUE provides 2-2.5 times faster recognition speed at the cost of a moderately increased error rate (1.5-2 times more errors). This property is available both for machine- and hand-printed texts. In the case of a hand-printed text (text type TT_Handprinted), a special recognition mode is used. On good print quality texts, ABBYY FineReader Engine makes an average of 1-2 errors per page, and such moderate increase in error rate can be easily tolerated in many cases, such as full text indexing with "fuzzy" searches, preliminary recognition, etc. *Note:* We do not recommend using this mode to recognize small image fragments (for example, fragments which consist of only one line or word) because the time advantage will be insignificant. By default, this property is FALSE.

Fine tuning		
LowResolutionMode	BOOL	Specifies whether a text on an image with low resolution is recognized. This property is useful when recognizing faxes, small prints, images with low resolution or bad print quality. By default, the value of this property is FALSE.
OneLinePerBlock	BOOL	This property set to TRUE tells ABBYY FineReader Engine to presume that the text in block to which the current RecognizerParams object belongs contains no more than one string. By default this property is FALSE.
OneWordPerLine	BOOL	This property set to TRUE tells ABBYY FineReader Engine to presume that no text line may contain more than one word, so the lines of text will be recognized as a single word. By default this property is FALSE.
ProhibitItalic	BOOL	This property set to TRUE tells ABBYY FineReader Engine not to recognize letters printed with italic -styled font. It is useful when a text with presumably no italic letters is recognized, in which case it may speed up the recognition. If there exist any italic letters on the image, and this property is TRUE, these letters will be recognized incorrectly. By default this property is FALSE.

ProhibitSubscript	BOOL	This property set to TRUE tells ABBYY FineReader Engine not to recognize subscript letters. It is useful when a text with presumably no subscripts is recognized, in which case it may speed up the recognition. If there exist any subscript letters on the image, and this property is TRUE, these letters will be recognized incorrectly. By default this property is FALSE.
ProhibitSuperscript	BOOL	This property set to TRUE tells ABBYY FineReader Engine not to recognize superscript letters. It is useful when a text with presumably no superscripts is recognized, in which case it may speed up the recognition. If there exist any superscript letters on the image, and this property is TRUE, these letters will be recognized incorrectly. By default this property is FALSE.
ProhibitHyphenation	BOOL	This property set to TRUE prohibits recognition of hyphenation from line to line. It is useful when a text with presumably no hyphenations is recognized, in which case it may speed up the recognition. If there exist any hyphenations in the recognized block, and this property is TRUE, the hyphenated words will be recognized incorrectly. By default this property is FALSE.
ProhibitInterblock-Hyphenation	BOOL	This property set to TRUE tells ABBYY FineReader Engine to presume that text from one block cannot be carried over to the next block. By default this property is FALSE.
CaseRecognitionMode	CaseRecognition-ModeEnum	This property specifies the mode of letter case recognition. Possible values are: CRM_AutoCase, CRM_SmallCase, CRM_CapitalCase. By default the value of this property is CRM_AutoCase, which corresponds to automatic case recognition.

Handprint recognition		
WritingStyle	WritingStyleEnum	Provides additional information about handprinted letters writing style. By default the value of this property is WS_Auto, which means that the writing style is automatically detected by FineReader Engine.

FieldMarkingType	FieldMarkingType-Enum	This property specifies the type of marking around letters (for example, underline, frame, box, etc.). This property is valid only for the handprint recognition. *Note:* For correct handprint recognition use CellsCount property that allows you to set the number of character cells for a recognized block. Possible values are: FMT_SimpleText, FMT_UnderlinedText, FMT_TextInFrame, FMT_GrayBoxes, FMT_CharBoxSeries, FMT_SimpleComb, FMT_CombInFrame, FMT_PartitionedFrame. By default the value of this property is FMT_SimpleText, which means the plain text.
CellsCount	int	Specifies the number of character cells for a recognized block. This property is valid only for the handprint recognition. It has a sense only for the field marking types (the FieldMarkingType property) that imply splitting the text in cells. Default value for this property is 1, but you should set the appropriate value to recognize the text correctly.

User patterns		
UseBuiltInPatterns	BOOL	This property set to TRUE means that ABBYY FineReader Engine will use its own built-in patterns for recognition. Patterns are files establishing relationship between character image and character itself. You may want to set this property to FALSE when you do not want to use standard ABBYY FineReader Engine patterns for character recognition, but user patterns only. This may be useful for recognition of text typed with decorative or nonstandard fonts. In this case it is better not to use ABBYY FineReader Engine built-in patterns, but use your own userdefined patterns trained for these fonts. A path to user-defined pattern file is stored in the UserPatternsFile property. If the UserPatternsFile property is empty the UseBuiltInPatterns property is ignored. See also Recognizing with Training. By default this property is TRUE.
UserPatternsFile	String	Contains the full path to a file of the user pattern used for recognition. If the value of this property is not empty, information from the user pattern file will be used during recognition. If the UseBuiltInPatterns property is FALSE, which means that standard ABBYY FineReader Engine patterns are not used during recognition, this property should contain a path to userdefined pattern file, as only information stored in it will be used. See also Recognizing with Training. By default this property stores an empty string.

Additional recognition information		
ErrorHighlightLevel	ErrorHighlight-LevelEnum	Specifies the level at which the ICharParams::IsSuspicious property is set to TRUE for a recognized character. The name of the property reflects the fact that the uncertain characters are highlighted with color in ABBYY FineReader. Possible values are: EHL_None, EHL_Scanty, EHL_Standard, EHL_Thorough, EHL_AllText. By default the value of this property is EHL_Standard.
ExactConfidence-Calculation	BOOL	If this property is TRUE, character and word confidence will be defined more accurately, but recognition speed may get slower. This property is automatically set to TRUE if the SaveCharacterRecognitionVariants or SaveWordRecognitionVariants property is TRUE. By default, this property is FALSE.
SaveCharacterRegions	BOOL	Specifies whether the exact characters regions (ICharParams::CharacterRegion) are saved. The default value is FALSE.
SaveCharacter-RecognitionVariants	BOOL	Specifies whether the variants of characters recognition are saved. The default value is FALSE.
SaveWordRecognition-Variants	BOOL	Specifies whether the variants of recognition of a word are saved. The default value is FALSE.

ATTENTION: Removed or renamed properties		
ErrorHiliteLevel	Renamed	The name of the property has been corrected, now it is ErrorHighlightLevel.
PossibleTextTypes TextType	Removed	The properties are obsolete and were deprecated in the previous version. Use the TextTypes property instead.
CJKTextDirection	Removed	Use the ITextOrientation::ReadingType property instead. If you set this property to TRT_Unknown, text direction will be detected automatically. By default, this property is TRT_Unknown as well.

Synthesis Parameters for Page

[SynthesisParamsForPage]

Main settings		
ParagraphExtraction-Mode	Paragraph-ExtractionMode-Enum	Specifies the mode of paragraph extraction. Possible values are: PEM_NormalExtraction, PEM_RoughExtraction, PEM_SingleLineParagraphsWithSpaceFormattin PEM_SingleLineParagraphsWithWordSeparationOnly. The default mode is PEM_NormalExtraction.
DetectFontFormatting-AtPageLevel	BOOL	If this property is set to TRUE, font parameters are detected at the stage of page synthesis. This property set to TRUE enables detection of subscripts, superscripts, italic-face type, small capital letters at the stage of page synthesis and allows you to set additional parameters using FontFormattingDetectionParams property. If this property is FALSE, the FontFormattingDetectionParams property is ignored. Important: With the default settings, ABBYY FineReader Engine 11 detects font parameters at the stage of document synthesis. If you set the value of this property to TRUE, you must turn off detection of font parameters during document synthesis. To do this, set the ISynthesisParamsForDocument::Detect FontFormatting property to FALSE. Detection of font parameters during page synthesis enables the program to speed up the subsequent document synthesis and decrease memory usage. However, the quality of font detection may deteriorate. This property is set to FALSE by default.

Color settings		
DetectBackgroundColor	ThreeState-PropertyValue-Enum	If this property is set to TSPV_Yes, the background color is detected during page synthesis. Possible values are: TSPV_No, TSPV_Yes, TSPV_Auto. This property is set to TSPV_Auto by default.
AllowGrayBackground-Color	ThreeState-PropertyValue-Enum	If this property is set to TSPV_Yes, the gray color is detected for background. Otherwise, background will be detected as black or white. The value of this property is taken into account only if the DetectBackgroundColor property is set to TSPV_Yes or TSPV_Auto. Possible values are: TSPV_No, TSPV_Yes, TSPV_Auto. The default value of this property is TSPV_Auto.
DetectTextColor	ThreeState-PropertyValue-Enum	If this property is set to TSPV_Yes, the text color is detected during page synthesis. Possible values are: TSPV_No, TSPV_Yes, TSPV_Auto. This property is set to TSPV_Auto by default.

CorrectDynamicRange	ThreeState-PropertyValue-Enum	If this property is set to TSPV_Yes, image colors will be corrected so that the background is white and the text is black, or vice versa, which improves image quality. Recognition, however, will slow down. We recommend using this property only if the DetectBackgroundColor and DetectTextColor properties are set to TSPV_Yes or TSPV_Auto. Possible values are: TSPV_No, TSPV_Yes, TSPV_Auto. This property is set to TSPV_Auto by default.

ATTENTION: Removed or renamed properties		
AllowGrayBackgroundColor AllowGrayTextColor CorrectDynamicRange DetectBackgroundColor DetectTextColor	Type changed	These properties now take values of the new ThreeStatePropertyValueEnum, which provides the opportunity not only to turn the setting on and off, but also make the selection automatically (TSPV_Auto constant). All these properties have TSPV_Auto value by default.

Synthesis Parameters for Document

[SynthesisParamsForDocument]

Main settings		
DetectDocument-Structure	BOOL	Specifies whether document structure detection should be performed while document synthesis. This property is set to TRUE by default.
DetectFontFormatting	BOOL	Specifies whether font formatting detection should be performed while document synthesis. If this property is set to FALSE, the FontFormattingDetectionParams section is ignored. Important: With the default settings, ABBYY FineReader Engine 11 detects font parameters at the stage of document synthesis. If you set the value of this property to FALSE, you must turn on detection of font parameters during page synthesis. To do this, set the DetectFontFormattingAtPageLevel property in the SynthesisParamsForPage section to TRUE. Detection of font parameters during page synthesis enables the program to speed up the subsequent document synthesis and decrease memory usage. However, the quality of font detection may deteriorate. This property is set to TRUE by default. *Note:* If your license does not support the Synthesis module, the default value of this property is FALSE.

Additional settings		
LowMemoryMode	BOOL	Specifies whether low memory mode should be used during document synthesis. If you set this property to TRUE, ABBYY FineReader Engine will try to use no more than 600 Mb of memory during document synthesis by loading less pages to memory simultaneously. However, document synthesis may slow down and the quality may slightly deteriorate. By default, the value of this property is FALSE.
PagePoolSize	int	Specifies how many pages may be loaded by document synthesis simultaneously. This property allows you to decrease memory usage. We recommend to use the value in range from 32 to 64. The more the value, the more speed of processing. However, for processing big documents it is not recommended to use the highest values of this property, as this may lead to an out of memory error. The value less than 5 is ignored. By default the value of this property is 64.

Font Formatting Detection Parameters

[FontFormattingDetectionParams]

Text decoration		
DetectBold	BOOL	If this property is set to TRUE, boldface is detected during synthesis. This property is set to TRUE by default.
DetectItalic	BOOL	If this property is set to TRUE, italic typeface is detected during synthesis. This property is set to TRUE by default.
DetectSubscriptsSuperscripts	BOOL	If this property is set to TRUE, subscript and superscript are detected during synthesis. This property is TRUE by default.
DetectUnderlineStrikeout	BOOL	If this property is set to TRUE, underline and strikeout are detected during synthesis. This property is set to TRUE by default.
DetectSmallCaps	BOOL	If this property is set to TRUE, small capital letters are detected during synthesis. This property is set to TRUE by default.

Font attributes		
DetectFontFamily	BOOL	If this property is set to TRUE, font name is detected during synthesis. This property is set to TRUE by default.
DetectFontSize	BOOL	If this property is set to TRUE, font size is detected during synthesis. This property is set to TRUE by default.

DetectSerifs	BOOL	If this property is set to TRUE, serif is detected during synthesis, i.e. if serif has been detected, serif typeface is selected to represent the recognized text. If this property is set to FALSE, serif is ignored. This means that the most suitable font (from both serif and sans serif typefaces) is selected to represent the recognized text, no matter whether the text is serif or sans serif. This property is set to TRUE by default.

Scaling and spacing		
DetectScaling	BOOL	If this property is set to TRUE, scaling is detected during synthesis. This property is TRUE by default.
DetectSpacing	BOOL	If this property is set to TRUE, spacing is detected during synthesis. This property is TRUE by default.
MonospaceDetectionMode	Monospace-DetectionMode-Enum	Specifies the mode of monospace font detection. Possible values are: MDM_Auto, MDM_NotMonospace, MDM_Monospace, MDM_Ignore. The default mode is MDM_Auto.

ATTENTION: Removed or renamed properties		
DetectMonospace	Removed	Use the MonospaceDetectionMode property instead. If you do not want monospaced typeface to be detected, set the MonospaceDetectionMode property to MDM_Ignore.

D Tesseract

D.1 Languages

The languages listed below are supported. Setting the correct language normally improves the result of the OCR process. The value that is to be passed to the switch -ocl is the three letter abbreviation.

Example: Set the Language to German

```
-ocl deu
```

English	**eng**
German	deu
French	fra
Italian	ita
Spanish	spa
Portuguese	por
Basque	eus
Dutch	nld

D.2 OCR Engine Parameters

This chapter lists supported key-value pairs. Values written in bold indicate the default value of this key. Neither the keys nor the values are case-sensitive. Currently there is only one setting supported.

Example: Enable automatic base line detection.

```
-ocp "ls=true"
```

`ls`	**FALSE**, TRUE	Automatically detect the base line of text. Setting this to true may improve the vertical alignment of text.